AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

FALL 1980

LECTURE NOTES IN ECONOMETRIC METHODS AND ANALYSIS

by

Joseph Havlicek, Jr.

Department of Agricultural Economics
and Department of Statistics
Virginia Polytechnic Institute
and State University

Contents

Page

Chapter 1: Nature of Econometrics

Introduction 1-1
Components of Applied Econometrics 1-1
Discipline in Econometrics 1-3
Assumptions About the Real World 1-5
Three Approaches to Analyzing Economic Data. 1-6

Chapter 2: Linear Regression

Introductory Remarks 2-1
Model Expression 2-2
Estimation of Regression Parameters. 2-5
Assumptions 2-6
Least Squares Estimators. 2-8
Properties of Least Squares Estimators. 2-14
Estimation From a Moment Viewpoint. 2-17
Some Measures of Variation. 2-23
Estimation From a Correlation Coefficient Viewpoint and Standardized Regression 2-28
Interval Estimation. 2-37
Confidence Limits for the Difference Between Two Regression Parameters 2-40
Confidence Limits for Predictions 2-41
Some Numerical Examples. 2-46
Estimation Using Correlation Coefficients 2-46
Point and Interval Estimation of Coefficients or Multiplicative Functions 2-53

Chapter 3: Testing Hypotheses About Linear Regression Parameters

Introduction 3-1
Fundamentals of Hypothesis Testing 3-1
Tests Concerned With Accounted for Variation. 3-3
Accounted for Variation and Deletion of Variables 3-5
Illustration of Testing Addition to Sum of Squares Regression. 3-7

Page

Testing Hypothesis About Single Equations 3-12

Testing Hypothesis About Pairs of Coefficients. 3-14

Testing Linear Combinations of Regression Coefficients 3-16

Orthogonal Comparisons. 3-18

Some Fundamentals of Linear Forms 3-23

Summary or Orthogonal Comparisons and Linear Forms. . . . 3-30

Newman-Kuels Test 3-30

Equality Between Coefficients of Two or More Relations 3-36

Chapter 4: Use of Zero-One Variables in Linear Regression

Introduction 4-1

Zero-One Variables as Intercept Shifters 4-3

Two or More Sets of Zero-One Variables as Intercept Shifters . 4-12

One Set of Zero-One Variables as Slope Shifters. 4-18

Two or More Sets of Zero-One Variables as Slope Shifters . . . 4-23

One Set of Zero-One Variables as Intercept and Slope Shifters. 4-27

Concluding Remarks About Zero-One Variables as Intercept and Slope Shifters 4-28

Matrix Representation of Zero-One Variables as Intercept and Slope Shifters 4-31

Zero-One Dependent Variable. 4-37

Discriminate Analysis. 4-40

Chapter 5: Autocorrelation and Serial Correlation

Introduction 5-1

Simple Illustration of Serial Correlation. 5-3

Effects of Serial Correlation. 5-6

Identification of Serial Correlation Problems. 5-7

Some Difficulties in Testing for Serial Correlation. 5-13

Alternative Approaches for Handling Serial Correlation Problems 5-22

Reformulate Model - Include Excluded Variables. 5-22

Interative Estimation of ρ. 5-23

Assume a Value of ρ 5-24

Incorporate Lagged Variables. 5-25

Use Knowledge of Variance-Covariance Matrix 5-26

Simultaneous Estimation of All Parameters 5-29

Page

Chapter 6: Generalized Least Squares

Introduction 6-1

Chapter 7: Heteroscedasticity

Introduction 7-1

Chapter 8: Errors in Variables

Introduction 8-1

Errors Only in the Endogenous Variable 8-1

Errors Only in the Predetermined Variables 8-3

Errors in Both Endogenous and Predetermined Variables. 8-5

Word of Caution About Measurement Errors 8-9

Chapter 9: Multicollinearity

Introduction 9-1

Consequences of Multicollinearity. 9-2

Identification of Multicollinearity Problems 9-3

Delete Variables. 9-4

Combined Variables. 9-5

Prior Information and Restrictions. 9-6

Transform Data. 9-6

Correct Inverse 9-7

Simulataneous Model 9-8

Chapter 10: Lagged Variables

Introduction 10-1

Underlying Motivation. 10-2

Nerlovian Lags 10-4

Expectations. 10-7

A Partial Adjustment Model. 10-9

Adaptive Expectation Model. 10-13

Some Potential Estimation Problems. 10-15

Koyck Types of Lags. 10-17

Caution About Using Distributed Lags 10-19

Polynomial Distributed Lags. 10-20

Page

Chapter 11: Models and Identification

Introduction 11-1
Structure and Economic Models. 11-1
Statistical Model and Variables. 11-2
Classification of Variables. 11-5
Role of Model in Determining Estimation Method 11-6
Completeness of a Model. 11-8
Reduced Form Equations 11-9
Identification 11-10
Simple Supply and Demand. 11-12
A Model With More Information in the Demand 11-17
A Model With More Information in the Supply 11-23
Just Identified Model 11-24
Over-Identified Model 11-28
Order and Rank Conditions for Identification 11-32
Order Condition 11-34
Rank Condition. 11-34
Other Information Which May Facilitate Identification. . . . 11-41
Independence of Disturbances. 11-42
Uncorrelated Disturbances 11-42
Limit to Size of Disturbance. 11-43
Variance of One Disturbance Less than the Variance of the Other 11-44

Chapter 12: Introduction to Simultaneous Linear Systems of Equations

Introduction 12-1
Full Information Methods 12-1
Limited Information Methods. 12-2
Properties of Estimators 12-3
Normalization of Equations 12-4
Least Squares Applied to Structured Equations. 12-7

Chapter 13: Two Stage Least Squares Estimation

Introduction 13-1
The Model and Assumptions. 13-1
A Cursory Look at Two Stage Least Squares. 13-2
The General Case 13-5
Testing and Interpretation 13-11

Page

Chapter 14: Limited Information Single Equation Estimation
Introduction 14-1
Least-Variance Ratio Criterion 14-1
Assumptions. 14-5
Estimation Procedure 14-5

Chapter 15: k Class, Double k Class and h Class Estimators. . . 15-1
Introduction 15-1
k Class Estimators 15-1
h Class Estimators 15-5
Double k Class Estimators. 15-6

Chapter 16: Instrumental Variables
Introduction 16-1
Choice of Instruments. 16-1
Just-Identified System 16-2
Over-Identified Systems. 16-4
Concluding Comment 16-5

Chapter 17: Three-Stage Least Squares
Introduction 17-1
Estimation Procedure 17-2

Chapter 18: Maximum Likelihood Estimation
Introduction 18-1
Likelihood and Probability 18-1
Likelihood Function. 18-2
Simple Linear Relations. 18-4
Illustration of Jacobians 18-6
Single Equation Likelihood Function 18-6
Multiple Variable Linear Relations 18-11
Simultaneous Linear Equations. 18-13

Page

Chapter 19: Derived Reduced Forms and Final Forms of Equations

Introduction 19-1

Difficulties Encountered in Interpreting Estimated Structural Coefficients. 19-1

Simultaneous Equations Model. 19-4

Analytically Derived Reduced Forms. 19-7

Final Form of Econometric Equation Systems 19-9

Impact Multipliers. 19-11

Interim Multipliers 19-11

Total Multipliers 19-12

Summary. 19-13

Chapter 20: Seemingly Unrelated Regressions

Introduction 20-1

Application of SUR Estimation. 20-2

Estimation Procedure 20-4

Estimating the Variance-Covariance Matrix. 20-8

An Application 20-11

Gains in Efficiency. 20-13

Chapter 21: Recursive Models. 21-1

Introduction 21-1

Foundations for Recursive Systems. 21-2

Empirical Recursive Models 21-6

Cournot and Walras. 21-9

Recursive and Simultaneous Systems Compared. 21-11

Summary and Conclusions. 21-15

Chapter 1: The Nature of Econometrics

Introduction

Econometrics is concerned with measurement of economic relationships. It is a science which attempts to describe an economy or part of an economy in terms of an equation or set of equations. There are both theoretical and empirical approaches to econometrics. The purely theoretical component consists of the development of a body of knowledge which tells us how to go about measuring economic relationships. The theory is developed on a fairly abstract or general basis, so that the results may be applied to any one of a variety of problems which may arise. Empirical procedures are direct applications of the methods of theoretical econometrics.

Econometricians attempt to express economic theories in mathematical terms in order to verify them by statistical methods and to measure the impact of a variable or set of variables, some but not necessarily all of which are economic variables, on one or a set of economic variables. Usually they are interested in this for one of two reasons. They may be interested in predicting future events arising from changes in certain economic variables or existing conditions. Or they may wish to advise about alternative economic policies to achieve certain desired results.

Components of Applied Econometrics

There are four major parts of applied or empirical econometrics. First, there is specification. Specification entails expressing an economic theory in mathematical terms. This is the model building activity.

A model is a set of mathematical expressions usually equations. In model building, one is always faced with the problem of what to leave out, equations and variables, if the model is to be manageable and yet useful given the data that are available.

The second component is estimation. Estimation entails "fitting" the model, in total or in part, to empirical data. The equations of the model are confronted with data. First the economic model has to be translated into a statistical model. This involves expressing the equations of the model in terms of statistical equations that is the relationships of the model which are not identities are expressed in a stochastic form. Each of the equations has to be expressed in a particular functional form and each of the variables in the equations has to specifically be defined in terms of the real world observations it represents. Then the usually scanty data available are used to obtain good and trustworthy estimates (numerical values) of the parameters (coefficients)of the statistical equations of the parameters (coefficients) of the statistical equations of the model.

The third component is verification of the model. Verification entails testing the model and the economic theory underlying the model. Criteria of success are established to accept or reject the economic theory being tested with the model and the particular set of data. Acceptance or rejection of the economic theory underlying the model refers to an acceptance or rejection of the applicability of the particular theory or theories used in terms of their ability to explain the observations of the particular economic phenomena given the particular economic phenomena given the particular statistical model and data used. In particular rejection simply means that the economic theory used did not satisfactorily describe the observed phenomena according to the criteria chosen. It is not possible to reject a theory

in an unqualified sense on the basis of an empirical application. Even if a given economic theory does not satisfactorily describe observations of economic variables in one situation it may still be a perfectly good theory for other samples of data and other situations.

The testing of a model in the verification aspect of econometrics involves testing via statistical methods, hypotheses about parameters or sets of parameters of the statistical equation. Conclusions and inferences are made about these parameters and the interpretation and conclusions about the economic theory are done in light of these statistical results. Generally tests are not performed with regard to hypotheses concerned directly with parameters of the non-statistical representation of the economic model. Both estimation and verification rely heavily on statistical methods and statistical theory.

The fourth component is prediction. This is what one is ultimately after in applied econometrics. One generally wants to determine and analyze what kind of changes are likely to occur if changes occur in certain variables or if certain relationships of the model experience shifts. Prediction may involve rearranging the model into a convenient shape and form, so that new information in predetermined variables can be fed into the model in order to obtain answers about the impact of these variables on endogenous variables in the system.

Disciplines in Econometrics

Econometrics is a combination of economics, mathematics and statistics. In this sense it is an interdisciplinary science. Let's briefly consider the role of each of these disciplines in econometric analysis.

Economic theory is fundamental to all aspects of econometrics. It establishes the boundaries within which we ask questions. The questions we pose and seek answers to are of an economic nature. Economic theory provides us with the hypotheses to be tested. In every piece of research we are concerned with some type of hypothesis which has evolved from some part of economic theory. From the micro theory of the firm evolve hypotheses about technical activities in terms of production functions and related marginal relationships, substitution among inputs in production, cost structures of firms and supply responses of individual firms and firms in an aggregate. From the micro theory of consumer behavior evolve hypotheses about marginal utilities and marginal rates of substitution, price response, income response, and demands for commodities for certain aggregates of consuming units. From various market oriented theories have evolved hypotheses about industry supplies, demand for inputs of industries, price behavior and industry cost structures. Business cycle theories have produced a large number of hypotheses without explicit reference to the theories of the firm, consumer behavior or such and these hypotheses have generally been in aggregative form referring to a whole industry, groups of industries, groups of households, etc. Certain macro theories have suggested hypotheses about consumption, savings, investment, employment, etc. One could continue listing various other theories such as monetary, growth, trade, and others and the hypotheses which have evolved from them but this list is meant only to illustrative and not exhaustive.

Mathematics is a tool both is economics and statistics. It provides a means of stating complex relationships in relatively simple

forms. It is a very useful "shorthand" in this sense. It permits us to write in a few equations what would take pages to explain narratively. Mathematics allows us to test for internal consistency and validity of conclusions derived from a set premises. The conclusions hold if the premises are valid and the argument is logical. Mathematics plays a similar role in statistics. It provides the foundation and theorectical basis for statistical procedures.

In econometrics, we are interested in the estimation and testing procedures of statistics. Many of these procedures involve operating with mathematical tools. Statistics provides us with a means of measuring from empirical data statistical representations of economic relationships and testing statistical hypotheses which have evolved from economic hypotheses. It provides the vehicle for making inductive inferences from a sample to a population. The use of statistics is particularly crucial in economics and other social sciences because there is little opportunity for controlled experiments. Generally we have to take the observed data as they have been generated by an economy or some part of an economy and draw inferences about the mechanism or structure which is believed to have generated the data. One reason why econometrics deserves separate treatment as a distinct body of knowledge is that the techniquea of measurement and testing developed are applied to non-experimental economic observations.

Assumptions About the Real World

For purposes of econometric analysis the world is assumed to have three essential characteristics. First, the economic system can be described by a system of equations which relate economic variables that

are consistent with economic theory. Second, the relationships are measurable. Third, the relationships are an autonomous set. The latter assumption means that the relationships are reasonably stable and are not going to change freely in structure. This doesn't mean that no change will ever occur but rather if a change does occur it can be explained by a limited number of economic variables. Furthermore, this means that a change may occur in the parameters of one equation of a system of equations without affecting the parameters of other equations in the system. For example, in a model consisting of a supply relationship and a demand relationship if for some reason one or more parameters of the supply relationship change the parameters of the demand relationships will remain unchanged if the relationships are autonomous. In other words a change in the behavioral response of suppliers doesn't change the behavioral response of demanders. The notion of autonomy is not an all or nothing proposition but rather a more or less proposition in terms of degrees of autonomy.

Three Approaches to Analyzing Economic Data

One approach which most of you are familiar with is to look at the time series in its own terms. You probably came across this approach in a course in prices, business statistics, basic statistical principles, or business cycle course. The approach attempts to decompose and classify the various types of fluctuations in a time series of data and identify the different classes with economically interpretable things. The four classes of fluctuations generally considered are trend, cycle, seasonal, and random. Frequently the tie between the classes of fluctuations or "wiggles" is loose and casual.

A second approach consists of extrapolation or forecasting. Emphasis is predicting what is going to happen with little emphasis on why it is going to happen. If there is considerable inertia in the economy under consideration the predictions may be successful without doing hardly any explaining of why things occurred the way they did. Much of the time the predictions may be good except when a turning point is reached.

A third approach is to develop and estimate statistical models based on economic theory which explain the time patterns of observable variables. This is the econometric approach. Emphasis is on explanation and understanding of behavior. It is different from the first approach in that each series involved is treated by itself but not independently of other series. Focus is on identifying causality and causal flows. Although interest in this approach may be in prediction and forecasting it differs from the second approach in terms of its stress on economic structural relationships and causality. The model may consist of one relationship or set of relationships which may be considered independently and handled statistically by single equation methods. Alternatively the model may consist of a simultaneous system of equations and estimation and analysis performed by one of the simultaneous equation procedures.

Chapter 2: Linear Regression

Introductory Remarks

In a number of instances a single equation representation of various economic phenomena is a suitable framework for estimation and analysis. Frequently such a representation entails a linear relationship in which the variation in a single endogenous variable is being accounted for by one or more exogenous or independent variables. The single exogenous variable relationship is a special case of the multiple exogenous variable regression. The relationships are specified as if causality was unidirectional from the exogenous variables to the endogenous variable even though in some cases there may be no intent to focus attention on causality per se. This is sometimes true about variables in models designed primarily for prediction purposes. On the other hand, models focusing on economic structure and the understanding of economic behavior are very much concerned with the directional flow of causality. In the latter types of models considerable emphasis is given to the underlying economic theory and other prior information which suggests the direction of causality.

The discussion of the linear regression model will focus on a number of topics which may be of interest to the applied economist. First, the linear model will be presented in compact matrix notation and the required assumptions for least squares estimators to be obtainable and have desirable properties will be discussed. This will be followed by a brief demonstration of the desirable properties of least squares estimaters when assumptions are fulfilled. The relationships among the underlying normal equations when expressed in general matrix notation, moments or sums of squares and cross-products and correlation coefficients will be pointed out.

Several sections will be devoted to point and interval estimation and testing of various types of hypotheses about individual parameters and groups of parameters. Finally, various common estimation and testing problems occur in empirical application including violation of assumptions will be discussed.

The emphasis is on the estimation aspects and application. The early sections contain a few proofs or possibly more accurately expressed as demonstration, many of which have been borrowed from Theil, Johnston, Goldberger and others. Most of the text is a blend of personal notes with texts of the individuals mentioned and others. Numerical examples are used liberally and consist mainly of performing various types of analyses and tests on a few selected sets of data so that the reader can more easily relate the various topics.

Model Expression

Let us start with a situation where an economic model or other type of model has been formulated and translated into a stochastic or statistical model. Also, let's suppose we have a set of observations or data which will be used to obtain numerical values of the coefficients of the statistical model. The set of data consists of n concurrent observations on one endogenous variable, Y, and k exogenous variables X_1, X_2, X_3, ... X_k,

$$
\begin{array}{llllll}
Y_1 & X_{11} & X_{12} & X_{13} & - - - & X_{1k} \\
Y_2 & X_{21} & X_{22} & X_{23} & - - - & X_{2k} \\
\cdot & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & \cdot & & \cdot \\
Y_i & Y_{i1} & X_{i2} & X_{i3} & - - - & X_{ik} \\
\cdot & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & \cdot & & \cdot \\
Y_n & X_{n1} & X_{n2} & X_{n3} & - - - & X_{nk}
\end{array} \qquad (2.1)
$$

We are interested in summarizing the pattern of the data by expressing the relationship between the X's and the Y in the form of a multiple variable linear stochastic equation such as

$$(2.2) \quad Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + --- + \beta_k X_{ik} + \varepsilon_i$$

$$(2.3) \quad Y_i = \sum_{j=0}^{k} \beta_j X_{ij} + \varepsilon_i$$

where: $i = 1, 2, 3, \ldots, n$

$j = 0, 1, 2, \ldots, k$

Y is the endogenous variable

$X_0 = 1$

$X_1, X_2, --- X_k$ are k exogenous variables

ε is the disturbance or error term whose values are not observable. The disturbance represents the combined effects of all variables neglected by the model and not included in the equation. β_0 is the intercept whose value is unknown. $B_1, \beta_2, --- \beta_k$ are parameters, the population regression coefficients, whose values are also unknown. In fact, attempting to obtain

numerical estimates of the parameters β_0, β_1, β_2, ---, β_k is what the estimation aspects of regression analysis are all about.

The observations presented in (2.1) may now be expressed in terms of the expressions presented in (2.2) and (2.3) in the following manner:

$$
\begin{aligned}
Y_1 &= \beta_0 X_{10} + \beta_1 X_{11} + \beta_2 X_{12} + - - - + \beta_k X_{1k} + \varepsilon_1 \\
Y_2 &= \beta_0 X_{20} + \beta_1 X_{21} + \beta_2 X_{22} + - - - + \beta_k X_{2k} + \varepsilon_2 \\
Y_3 &= \beta_0 X_{30} + \beta_1 X_{31} + \beta_2 X_{32} + - - - + \beta_k X_{3k} + \varepsilon_3 \qquad (2.4) \\
&\;\;\vdots \\
Y_i &= \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + - - - + \beta_k X_{ik} + \varepsilon_i \\
&\;\;\vdots \\
Y_n &= \beta_0 X_{n0} + \beta_1 X_{n1} + \beta_2 X_{n2} + - - - + \beta_k X_{nk} + \varepsilon_n
\end{aligned}
$$

Remember that the elements of $X_{10} - - - X_{n0}$ are each equal to one.

The n equations in (2.4) may also be expressed in matrix notation as

$$Y = X \beta + \varepsilon \qquad (2.5)$$

where:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & X_{23} & \cdots & X_{2k} \\ 1 & X_{31} & X_{32} & X_{33} & \cdots & X_{3k} \\ . & . & . & . & & . \\ . & . & . & . & & . \\ . & . & . & . & & . \\ 1 & X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nk} \end{bmatrix} \tag{2.6}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ , \\ . \\ \beta_k \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

Y is n x 1

X is n x (k+1)

β is (k+1) x 1

ε is n x 1

Estimation of Regression Parameters

There are many ways in which the numerical estimates of the β's could be chosen. One of the more popular methods is based on the least squares criterion. The least squares method involves fitting the multiple regression equation, which is a hyperplane, to the observations in such a way that the

sum of squares of the residuals is a minimum. No other set of numerical estimates of the β's can yield a smaller sum of squares of the residuals. The residuals are the numerical estimates of the ε's. When certain assumptions are met, least squares estimators have the desirable properties of being best, linear, and unbiased estimators.

Assumptions

There are five major assumptions underlying classical linear regression models in order for the estimators to be best, linear and unbiased. An additional assumption about the distribution of the disturbance has to be made if tests of hypotheses are to be performed.

(A.1) $Y = X\beta + \varepsilon$

This assumption states that the maintained hypothesis is linear in the sense that Y observations are linear functions of the X observations and the ε's and the parameters are linear functions of the Y observations. The exogenous variables may be of various powers.

(A.2) $E(\varepsilon) = 0.$

This assumption states that the expected value of the disturbance is zero. Hence, the mean value of the disturbance is zero. One way of thinking about this is that the influences of all the variables excluded from the model and not considered are expected to cancel each other out on the average and have no influence on the endogenous variable.

(A.3) $E(\varepsilon\ \varepsilon) = o^2 I$

This assumption really encompasses two assumptions. First, it states that $E(\varepsilon_i^2) = o^2$ for all i which indicates the ε_i have a constant variance. This is the homoscedasticity property of variances. This means the diagonal

elements of the matrix $\sigma^2 I$ are all the same. In fact $\sigma^2 I$ is a scalar matrix because of the second assumption embodied in (A.3). The off-diagonal terms are zero and this occurs because $E(\epsilon_i \epsilon_r) = 0$ for $i \neq r$ which means that the ϵ_i's (disturbances) are pairwise uncorrelated in the particular sample being considered.

(A.4) X is a n x (1+k) matrix of fixed numbers

This assumption states that the exogenous variables are nonstochastic or fixed in repeated samples. Hence, for a specified set of X's the sole variation in Y is variation in ϵ. This implies that X and ϵ are independent and hence $E(X\epsilon) = E(X) \cdot E(\epsilon) = E(X) \cdot 0 = 0$ because of (A.2)

(A.5) Rank of X = $1+k \leq n$

This assumption also consists of two parts which could be considered two assumptions. First, it says the rank of X is equal to the number of exogenous variables, including the "dummy" variable which takes on only the value of one, which means there are no exact linear relationships among the exogenous variables. The number of exogenous variables including the "dummy" variable of the intercept is also the number of parameters to be estimated. The second part of this assumption $(1+k) \leq n$ simply indicates the number of parameters to be estimated has to be equal to or less than the number of observations. If the equality holds, the fit of the regression equation shall be perfect because the regression surface shall be forced through every single observation and there shall be no residual of any type. If (k+1) were greater than n there would not be adequate information in sample observations to provide information about all of the parameters. Another way of saying this is that more degrees of freedom would be required than are available.

Finally if tests of significance are to be performed than the additional assumption required is:

(A.6) ε are normally distributed.

This assumption is required so that probability distributions with known parameters and thus properties can be used in specifying and evaluating tests of hypotheses. Note that since X is considered a set of fixed numbers and the β is a set of constants (parameters) the Y takes on the same assumed type of distribution as ε. Both Y and ε have the same variance σ^2, but the mean of Y is μ_Y while the mean of ε is $E(\varepsilon) = 0$. Under this assumption Y is $N(\mu_Y, \sigma^2)$ and ε is $N(0,\sigma^2)$.

Least-Square Estimators

In (2.5) a linear regression model with K exogenous variables was expressed in matrix notation. The specific model considered was

$$Y = X\beta + \varepsilon$$

Now let us consider the estimating form of this model

(2.7) $Y = Xb + e$

where: Y is a n x 1 vector of observations of the endogenous variable

X is a n x (k+1) matrix of observation of the exogenous variables[1/]

b is a (k+1) x 1 vector of estimators

e is a n x 1 vector of residuals of $(Y - Xb)$

There are two rather critical differences between (2.5) and (2.7). First, the vector of β parameters in (2.5) whose values are unknown is replaced

---

[1/] Remember, there are k exogenous explanatory variables plus the dummy variable which takes on only values of one; hence the dimensions.

in (2.7) by the b estimators. Second, the unobservable disturbances, ε, in (2.5) are replaced with the observable residuals, e, in (2.7).

The least squares principle is one criterion which may be used to obtain the b estimators. The minimization of the sum of squared residuals is the basis for this criterion. Considering a sample size of n, the sum of squared residuals is

$$(2.8)\quad e'e = (Y - Xb)'(Y - Xb) = \sum_{i=1}^{n} e_i^2$$
$$= (Y' - b'X')(Y - Xb)$$
$$= Y'Y - Y'Xb - b'X'Y + b'X'Xb$$

Note that Y'Xb is a scalar and thus equal to its transpose. Hence, $(Y'Xb)' = b'X'Y$ which permits combining Y'Xb and b'X'Y.

In order to minimize the sum of squared residuals the gradient is obtained and set equal to zero as follows:

$$(2.9)\quad \frac{\partial e'e}{\partial b} = -2\ X'Y + 2X'Xb = 0 \ \underline{2/}$$

and

$$(2.10)\quad X'Xb = X'Y$$

$$(2.11)\quad (X'X)^{-1} X'Xb = (X'X)^{-1} X'Y$$

$$(2.12)\quad Ib = (X'X)^{-1} X'Y$$

$$(2.13)\quad b = (X'X)^{-1}X'Y$$

$\underline{2/}$ Vector and matrix differentiation applied to a quadratic form was used to obtain (2.9). Suppose we have a quadratic form X'AX. $\frac{(X'AX)}{\partial X} = (A + A')X$ and if A is symmetric as is the case with X'X then $A' = A$ and $(A + A')X = 2AX = 2A'X$. Alternatively, the quadratic form would be expanded and then partially differentiated term by term.

Hence, (2.13) are the leasts squares estimators which are the recipes for obtaining from observed data numerical estimates of the regression parameters.

Now let us attempt to develop the covariance matrix of the least squares coefficient estimators. The sampling error is

$$(2.14)\quad b - \beta = (X'X)^{-1} X'Y - \beta$$

but since $Y = X\beta + \varepsilon$ then

$$(2.15)\quad \begin{aligned} b - \beta &= (X'X)^{-1} X' (X\beta+\varepsilon) - \beta \\ &= (X'X)^{-1} X'X\beta + (X'X)^{-1} X'\varepsilon - \beta \\ &= \beta + (X'X)^{-1} X'\varepsilon - \beta \\ &= (X'X)^{-1} X'\varepsilon \end{aligned}$$

Note that since the elements of $(X'X)^{-1} X'$ are constant that the sampling error $(b - \beta)$ is a linear function of ε, the disturbance. Also, $(b - \beta)$ is a similar function of the disturbance as b is of the endogenous variable.

Now,

$$\begin{aligned} (b - \beta)(b - \beta)' &= [(X'X)^{-1} X' \varepsilon][(X'X)^{-1} X'\varepsilon]' \\ &= [(X'X)^{-1} X' \varepsilon][\varepsilon' X (X'X)^{-1}] \\ &= (X'X)^{-1} X' \varepsilon\varepsilon' X(X'X)^{-1} \end{aligned}$$

The covariance matrix is the expected value of this expression given that the least squares estimators are unbiased.[3/]

$$(2.16)\quad V(b) = E[b - \beta](b - \beta)'] = E[(X'X)^{-1} X' \varepsilon\varepsilon' X (X'X)^{-1}]$$

but since the elements of $(X'X)^{-1} X'$ and $X(X'X)^{-1}$ are fixed or constant

$$(2.17)\quad \begin{aligned} V(b) &= (X'X)^{-1} X' E(\varepsilon \varepsilon') X(X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} X'X (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} \end{aligned}$$

There are two things to note about the covariance matrix of b. First, it entails, the same inverse matrix as do the least squares estimators,

---

[3/] The unbiasedness property will be proved in a subsequent section.

b. Second, the covariance matrix is in terms of the variance parameter, σ^2. Usually the true variance is unknown and an estimate of the variance is needed. One way of handling this is to substitute for the unobservable disturbance, $\varepsilon = Y - X\beta$, the observable residuals, $e = Y - Xb$, and utilize the residual variance as an estimator of σ^2.

The estimator of the residual variance associated with a least squares regression equation is:

$$(2.19) \quad s^2 = \frac{e'e}{n-k-1} = \frac{\sum_{i=1}^{n} e_1^2}{n-k-1} = \frac{\text{sum of squared residuals}}{\text{degrees of freedom}}$$

Intuitively, it is not difficult to see why this estimator appears appropriate and makes sense. In the least squares framework the sum of squared residuals is a measure of the variation in the endogenous variable not accounted for by the exogenous variables, X, included in the regression equation and when divided by the appropriate degrees of freedom yields an estimate of residual variance. Furthermore, intuitively n-k-1 degrees of freedom appear appropriate in that this is simply the number of observations of the variables less the number of restrictions imposed on the given sample of n observations in terms of the freedom in which the variables can assume values. There are k parameters of variables plus an intercept to be estimated which imposes k + 1 restrictions and hence k + 1 degrees of freedom are "used up".

To show that s^2 is the least squares estimator of residual variance and that n - k - 1 are the appropriate degrees of freedom involves some tedious but not extremely difficult matrix manipulations. First, let's consider the alternative forms of the least squares residual which are readily amenable to matrix manipulation.

$$(2.19)\quad e = Y - Xb = Y - X(X'X)^{-1}X'Y = \left[I - X(X'X)^{-1}X'\right]Y \text{ [4/]}$$

where $\left[I - X(X'X)^{-1}X'\right]$ is symmetric and idempotent.[5/] Now substituting $Y = X\beta + \varepsilon$ into (2.19) yields:

$$\begin{aligned}(2.20)\quad e &= \left[I - X(X'X)^{-1}X'\right]Y = \left[I - X(X'X)^{-1}X'\right](X\beta + \varepsilon)\\ &= \left[I - X(X'X)^{-1}X'\right]X\beta + \left[I - X(X'X)^{-1}X'\right]\varepsilon\\ &= X\beta - X(X'X)^{-1}X'X\beta + \left[I - X(X'X)^{-1}X'\right]\varepsilon\\ &= X\beta - X\beta + \left[I - X(X'X)^{-1}X'\right]\varepsilon\\ &= \left[I - X(X'X)^{-1}X'\right]\varepsilon\end{aligned}$$

Note that the least squares residual vector is a similar linear function of the disturbance, ε, as it is of the endogenous variable, Y. The sum of squared residuals is:

[4/] Any n x 1 vector is equal to the product of itself premultiplied by an n x n identity matrix, i.e., $Y = I_n Y$.

[5/] An idempotent matrix is a square matrix which satisfies the condition that $A = A^2$. In this case

$$\begin{aligned}\left[I - X(X'X)^{-1}X'\right]\left[I - X(X'X)^{-1}X'\right] &= I^2 - IX(X'X)^{-1}X'I\\ &\quad + X(X'X)^{-1}X'\;X(X'X)^{-1}X'\\ &= I - X(X'X)^{-1}X' = X(X'X)^{-1}X'\\ &\quad + X(X'X)^{-1}X'\\ &= I - X(X'X)^{-1}X'\end{aligned}$$

also, $\left[I - X(X'X)^{-1}X'\right]' = I' - \left[X(X'X)^{-1}X'\right] = \left[I - X(X'X)^{-1}X'\right]$ since $X'X$ and $(X'X)^{-1}$ are symmetric metracies and $\left[(X'X)^{-1}\right]' = (X'X)^{-1}$.

$$(2.22)\quad e'e = \varepsilon' \left[I - X(X'X)^{-1} X'\right]\varepsilon = \operatorname{tr}\{\varepsilon' \left[I - X(X'X)^{-1} X'\right]\varepsilon\} \text{ 6/}$$

e'e may be conveniently expressed in terms of a trace since $\varepsilon' \left[I - X(X'X)^{-1} X'\right]\varepsilon$ is a scalar and thus equal to its own trace.7/ The trace is a linear function and since the expected value of a trace is the trace of the expected value, taking the expectation of the sum of squared residuals yields:

$$(2.23)\quad E(e'e) = E\{e' \left[I - X(X'X)^{-1} X'\right]\varepsilon\} = E \operatorname{tr}\{\varepsilon' \left[I - X(X'X)^{-1} X'\right]\varepsilon\}$$

$$= E \operatorname{tr}\{\left[I - X(X'X)^{-1} X'\right] \varepsilon\varepsilon'\} = \operatorname{tr} E\{\left[I - X(X'X)^{-1} X'\right]\varepsilon\varepsilon'\}$$

$$= \operatorname{tr}\{\left[I - X(X'X)^{-1} X'\right] E(\varepsilon\varepsilon')\}$$

$$= \operatorname{tr}\{\left[I - X(X'X)^{-1} X'\right]\sigma^2 I\}$$

$$= \operatorname{tr}\{\left[I - X(X'X)^{-1} X'\right] \sigma^2\}$$

$$= \{\operatorname{tr}\left[I - X(X'X)^{-1}\right]\} \sigma^2$$

$$= \{\operatorname{tr}[I_n] - \operatorname{tr}[(X'X)(X'X)^{-1}]\}\sigma^2 \text{ 8/}$$

$$= \{\operatorname{tr}[I_n] - \operatorname{tr}[I_{k+1}]\} \sigma^2$$

$$= [n - (k-1)] \sigma^2 = (n-k-1) \sigma^2$$

6/ The trace of a square matrix is the sum of its diagonal elements, i.e., $\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$ where A is n x n.

7/ It may not be obvious that $\varepsilon' [I - X(X'X)^{-1} X']\varepsilon$ is a scalar, however, consider the size of each of the matrices and the order of multiplication.-

ε' is 1 x n
I is n x n
X is n x (k+1)
$(X'X)^{-1}$ is (k+1) x (k+1)
X' is (k+1) x n
ε is n x 1.

The result is a 1 x 1.

8/ This particular step is based on a basic property of traces of products of matrices. If A, B, and C are three matrices conformable to multiplication then tr (ABC) = tr (BCA) = tr (CAB). In this case A = X, B = $(X'X)^{-1}$, and C = X' and tr $[X(X'X)^{-1} X']$ = tr (ABC) = tr (CAB) = tr $[X'X (X'X)^{-1}]$. Also, this operation yields an identity matrix and the order of the identity matrix is its trace. The orders of the product matrices resulting from the various multiplication will not be the same unless all the component matrices of the product are square and all have the same order.

Hence, an unbiased estimator of the disturbance variance is obtained by dividing the error sum of squares or sum of squared residual by n - k - 1. Thus, the unbiased estimator of σ^2 is:

$$(2.24) \quad s^2 = \frac{e'e}{n-k-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1} = \frac{SSE}{n-k-1}$$

where SSE denotes errorssum of squares. Combining this information with (2.17) it directly follows that an unbiased estimator of the covariance matrix of b is:

$$(2.25) \quad S(b) = s^2 (X'X)^{-1}$$

The estimators of variances are the elements of the main diagonal while off diagonl elements are covariances.

Properties of Least Squares Estimators

We went through some tedious and messy matrix algebra to show that the estimator of the disturbance variance and the estimators of the variances and covariances of the regression coefficidnts are unbiased, but nothing has been said about the properties of the estimators of the regression coefficients. If assumptions (A.1) through (A.5) are met then the estimators have the desirable properties of being best, linear, and unbiased estimators (BLUE).

One of the easier to show properties is that b is linear. Linearity here refers to b being a linear function of Y, the endogenous variable. This can be done by considering the specification of b given in (2.13) which is

$$b = (X'X)^{-1} X'Y$$

In assumption (A.4), X was considered a matrix of fixed numbers, thus $(X'X)^{-1}$ X' consists of constant elements and b is linear in Y.

The estimators is unbiased if for any arbitrary sample size the expected value of the estimator is the parameter being estimated. To show this, let us take the expectation of b.

$$(2.26) \quad E(b) = E[(X'X)^{-1} [X'Y] = (X'X)^{-1} X' E(Y)$$

But, the expectation of Y is

$$(2.27)\quad E(Y) = E[X\beta + \varepsilon] = X\beta + E(\varepsilon) = X\beta$$

Thus,

$$(2.28)\quad E(b) = (X'X)^{-1} X' (X\beta) = (X'X)^{-1} (X'X)\beta = \beta$$

Hence, b is an unbiased estimator of β.

The "bestness" property is somewhat more difficult to prove. It entails showing for any arbitrary sample size that there is no other linear unbiased estimator that has a smaller variance than the least squares estimator. In a rigorous manner Theil states that if the assumptions (A.1) through (A.5) are not violated the least squares estimator b is best linear unbiased in the sense that any other estimator of β which is also linear in the vector Y and unbiased has a covariance matrix which exceeds that of b by a positive semi-definite matrix.[9/]

We have already shown that $b = (X'X)^{-1} X'Y$ is an unbiased estimator of β and is a linear function of Y. Also, we have shown that the covariance matrix of b is $V(b) = \sigma^2 (X'X)^{-1}$ which in the least squares framework can be estimated by $s^2 (X'X)^{-1}$. Now let us consider any other linear estimator of β, say $\overline{\beta}$, which can be expressed as

$$(2.29)\quad \overline{\beta} = A'Y$$

and A' can always be expressed as

$$(2.30)\quad A' = [(X'X)^{-1} X' + D']$$

where D' is a (k+1) x n nonstochastic matrix all of whose elements are not zero. For $\overline{\beta}$ to be unbiased $A'X = I + D'X = I$ which means

$$(2.31)\quad D'X = 0$$

[9/] For additional details see Chapter 3 of Henri Theil, Principles of Econometrics, 1971.

Under this condition and by substituting (2.30) into (2.29) the covariance matrix of $\bar{\beta}$ is

$$(2.32)\quad V(\bar{\beta}) = E[(\bar{\beta} - \beta)(\bar{\beta} - \beta)'] = E[(A'\epsilon)(A'\epsilon)']$$
$$= E[A'\epsilon\epsilon'A] = \sigma^2 A'A$$

where

$$(2.33)\quad A'A = [(X'X)^{-1}X' + D'][(X'X)^{-1}X' + D']'$$
$$= [(X'X)^{-1}X' + D'][X(X'X)^{-1} + D]$$
$$= (X'X)^{-1}X'X(X'X)^{-1} + (X'X)^{-1}X'D + D'X(X'X)^{-1} + D'D$$
$$= (X'X)^{-1} + D'D$$

$(X'X)^{-1}X'D$ and $D'X(X'X)^{-1}$ both vanish because of (2.31). Hence,

$$(2.34)\quad V(\bar{\beta}) = \sigma^2[(X'X)^{-1} + D'D] = \sigma^2(X'X)^{-1} + \sigma^2 D'D$$

where D'D is a positive semi-definite matrix.[10/] Thus, any linear estimator $\bar{\beta}$ which is unbiased, i.e., satisfies (2.31) will have variances of the estimators which exceed the variances of the least squares estimators by the amount of the diagonal elements of D'D. Covariances will differ by the amounts of the corresponding off-diagonal elements of D'D. Hence the least squares estimators are the best of all possible linear unbiased estimators.

[10/] A matrix premultiplied by its own transpose yields a product matrix with main diagonal elements which are a sum of squares and as long as we are dealing with real numbers the main diagonal elements are non-negative. Furthermore, D'D is square and symmetric. One can show it is symmetric by simply transposing D'D; (D'D)' = D'D

To show that D'D matrix is positive semi-definite, let us express it as the coefficient matrix of a quadratic form

$$F = Z'(D'D)Z$$

where

Z' is 1 x (k+1)
D' is (k+1) x n
D is n x (k+1)
(D'D) is (k+1) x (k+1)
Z is (k+1) x 1

If (D'D) is positive semi-definite then F is non-genative (≥ 0) for every Z and there exists points $Z \neq 0$ for which $Z'(D'D)Z = 0$. Note that F is a quadratic equation and in terms of matrices it is a 1 x 1. Now expressing F as

$$F = (Z'D')(DZ) = (DZ)'(DZ)$$

we note that Z'D' is 1 x n

Estimation From a Moment Viewpoint

Much of what has been set forth thus far has been theoretical in nature with focus on showing why the least squares estimators are what they are and showing that they do possess some desirable properties if certain key assumptions are fulfilled. Now we shall turn to developing more of an applied framework which directly will yield computational formuli for obtaining numerical values of the least squares estimators. Then some of the fundamental and major tests of hypotheses about parameters are presented.

As a point of departure, let us more closely examine the X'X matrix which plays such a key role in least squares estimation. The X matrix presented in (2.6) is a n x (k+1) consisting of n observations on each of k exogenous variables and a dummy variable, X_{10}, which only takes on values of one. Hence,

$$X = \begin{bmatrix} X_{10} & X_{11} & X_{12} & \cdots & X_{1k} \\ X_{20} & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ X_{n0} & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}$$

$$X'X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & & \vdots \\ X_{1k} & X_{2k} & \cdots & X_{nk} \end{bmatrix} \cdot \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}$$

$$(2.35)\quad X'X = \begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & \cdots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^2 & \sum_{i=1}^{n} X_{i1}X_{i2} & \cdots & \sum_{i=1}^{n} X_{i1}X_{ik} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i2}X_{i1} & \sum_{i=1}^{n} X_{i2}^2 & \cdots & \sum_{i=1}^{n} X_{i2}X_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{ik}X_{i1} & \sum_{i=1}^{n} X_{ik}X_{i2} & \cdots & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix}$$

The cross moments of the exogenous variables with the endogenous variable unadjusted for means (moments about the origin) are:

$$(2.36)\quad X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & & \vdots \\ X_{1k} & X_{2k} & \cdots & X_{nk} \end{bmatrix} \cdot \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{i1} Y_i \\ \vdots \\ \sum_{i=1}^{n} X_{ik} Y_i \end{bmatrix}$$

If the subscripts of X' in (2.36) appear peculiar remember that the columns of X' are the corresponding rows of X.

The normal equations of least squares estimators presented in matrices of moments unadjusted for means are:

$$(2.37)\qquad \begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & \cdots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^2 & \sum_{i=1}^{n} X_{i1}X_{i2} & & \sum_{i=1}^{n} X_{i1}X_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{ik}X_{i1} & \sum_{i=1}^{n} X_{ik}X_{i2} & & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{i1}Y_i \\ \vdots \\ \sum_{i=1}^{n} X_{ik}Y_i \end{bmatrix}$$

Numerical values of all elements in X'X and X'Y are obtainable from the observed data. If the inverse of X'X is obtained, (2.37) may be solved for estimates of the elements of the b vector. $(X'X)^{-1}$ entails elements which permit obtaining the variances and covariances of all estimated coefficients including the intercept, b_0. If tests of hypothesis about the intercept or interval estimates for the intercept are relevant to the problem being investigated, then the unadjusted moments form of normal equations are appropriate.

Alternatively the normal equations may be expressed in terms of corrected moments or moments adjusted for means (moments about the means). To show the normal equations expressed in terms of adjusted moments (2.37) may be directly solved for b_0 and this expression for b_0 substituted into each of the k remaining normal equations. The normal equations given by (2.37) written out are:

$$
\begin{aligned}
&nb_0 + b_1 \sum_{i=1}^{n} X_{i1} + b_2 \sum_{i=1}^{n} X_{i2} + \cdots + b_k \sum_{i=1}^{n} X_{ik} = \sum_{i=1}^{n} Y_i \\
&b_0 \sum_{i=1}^{n} X_{i1} + b_1 \sum_{i=1}^{n} X_{i1}^2 + b_2 \sum_{i=1}^{n} X_{i1} X_{i2} + \cdots + b_k \sum_{i=1}^{n} X_{i1}X_{ik} = \sum_{i=1}^{n} X_{i1} Y_i \\
&\quad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
&b_0 \sum_{i=1}^{n} X_{ik} + b_1 \sum_{i=1}^{n} X_{ik}X_{i1} + b_2 \sum_{i=1}^{n} X_{ik}X_{i2} + \cdots + b_k \sum_{i=1}^{n} X_{ik}^2 = \sum_{i=1}^{n} X_{ik}Y_i
\end{aligned}
\tag{2.38}
$$

Solving the first equation of (2.38) for b_0 yields

$$
b_0 = \frac{\sum_{i=1}^{n} Y_i}{n} - b_1 \frac{\sum_{i=1}^{n} X_{i1}}{n} - b_2 \frac{\sum_{i=1}^{n} X_{i2}}{n} - \cdots - b_k \frac{\sum_{i=1}^{n} X_{ik}}{n} \tag{2.39}
$$

For illustrative purposes substituting (2.39) into the second equation of (2.38) yields

$$
\begin{aligned}
&\left[\frac{\sum_{i=1}^{n} Y_i}{n} - b_1 \frac{\sum_{i=1}^{n} X_{i1}}{n} - b_2 \frac{\sum_{i=1}^{n} X_{i2}}{n} - \cdots - b_k \frac{\sum_{i=1}^{n} X_{ik}}{n}\right] \sum_{i=1}^{n} X_{i1} \\
&\quad + b_1 \sum_{i=1}^{n} X_{i1}^2 + b_2 \sum_{i=1}^{n} X_i X_{i2} + \cdots + b_k \sum_{i=1}^{n} X_{i1}X_{ik} = \sum_{i=1}^{n} X_{i1}Y_i \\
&= b_1 \left[\sum_{i=1}^{n} X_{i1}^2 - \frac{\left(\sum_{i=1}^{n} X_{i1}\right)\left(\sum_{i=1}^{n} X_{i1}\right)}{n}\right] + b_2 \left[\sum_{i=1}^{n} Y_{i1}X_{i2} - \frac{\left(\sum_{i=1}^{n} X_{i1}\right)\left(\sum_{i=1}^{n} X_{i2}\right)}{n}\right] \\
&\cdots + b_k \left[\sum_{i=1}^{n} X_{i1}X_{ik} - \frac{\left(\sum_{i=1}^{n} X_{i1}\right)\left(\sum_{i=1}^{n} X_{ik}\right)}{n}\right] = \sum_{i=1}^{n} X_{i1}Y_i - \frac{\left(\sum_{i=1}^{n} X_{i1}\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}
\end{aligned}
$$

Note that $\sum_{i=1}^{n} X_{ik}^2 - \frac{\left(\sum_{i=1}^{n} X_{ik}\right)^2}{n} = \sum_{i=1}^{n} (X_{ik} - \bar{X}_k)^2 = \sum_{i=1}^{n} x_{ik}^2$

and $\sum_{i=1}^{n} X_{ik}X_{ij} - \frac{\left(\sum_{i=1}^{n} X_{ik}\right)\left(\sum_{i=1}^{n} X_{ij}\right)}{n} = \sum_{i=1}^{n} (X_{ik} - \bar{X}_k)(X_{ij} - \bar{X}_j) = \sum_{i=1}^{n} x_{ik} x_{ij}$

Small letters are being used to denote deviations from the means and this convention will be adopted throughout the text. Rewriting (2.40) in this notation we have:

$$(2.41)\quad b_1 \sum_{i=1}^{n} x_{i1}^2 + b_2 \sum_{i=1}^{n} x_{i1}x_{i2} + \cdots + b_k \sum_{i=1}^{n} x_{i1}x_{ik} = \sum_{i=1}^{n} x_{i1}y_i$$

It is rather straight forward that each of the normal equations may be rewritten similarly to (2.41). The normal equations written in terms of adjusted moments and in matrix form are:

$$(2.42)\quad \begin{bmatrix} \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \cdots & \sum x_{i1}x_{ik} \\ \sum x_{i2}x_{i1} & \sum x_{i2}^2 & \cdots & \sum x_{i2}x_{ik} \\ \vdots & \vdots & & \vdots \\ \sum x_{ik}x_{i1} & \sum x_{ik}x_{i2} & \cdots & \sum x_{ik}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \sum x_{i1}y_i \\ \sum x_{i2}y_i \\ \vdots \\ \sum x_{ik}y_i \end{bmatrix}$$

If the inverse of [x'x] is obtained, the equations given in (2.42) may be solved for $b_1, b_2, \cdots, b_k$. b_0 is obtained from (2.39).

There are a few aspects of the normal equations expressed in terms of adjusted moments which are worthy of mention. First, notice that the set of least squares estimators of the coefficients may be obtained either from moments adjusted for means or moments unadjusted for means. Second, the matrix which has to be inverted when the moments are expressed about the means is a k x k as compared to a (k+1) x (k+1) when the X'X matrix is used. In the age of electronic computers this is not as important, but when calculations are

being done by hand reducing the matrix to be inverted by one row and one column may save a tremendous amount of work. Third, and probably most important, $\left[X'X\right]^{-1}$ does not contain elements which would permit obtaining the variance of the intercept nor covariances involving the intercept. Hence, if testing hypotheses about the intercept or obtaining interval estimates for the intercept are desired, the procedure involving the adjusted moments is not the one to use.

Again consider the normal equations in terms of unadjusted moments as expressed in (2.37) and obtain a solution for the b vector. Let $C = (X'X)^{-1}$ with elements c_{ij} where i denotes rows and j the columns of C. The estimators, b_i are obtained by

$$(X'X)^{-1}\ (X'X)\ b = (X'X)^{-1}\ X'Y$$

$$b = (X'X)^{-1}\ X'Y$$

$$b = C\ X'Y$$

(2.43)

$$C = \begin{bmatrix} c_{00} & c_{01} & c_{02} & \cdots & c_{0k} \\ c_{10} & c_{11} & c_{12} & \cdots & c_{1k} \\ c_{20} & c_{21} & c_{22} & \cdots & c_{2k} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ c_{k0} & c_{k1} & c_{k2} & & c_{kk} \end{bmatrix}$$

and

$$(2.44)\qquad \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_k \end{bmatrix} = \begin{bmatrix} c_{00} & c_{01} & c_{02} & \cdots & c_{0k} \\ c_{10} & c_{11} & c_{12} & \cdots & c_{1k} \\ c_{20} & c_{21} & c_{22} & \cdots & c_{2k} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ c_{k0} & c_{k1} & c_{k2} & \cdots & c_{kk} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{i1} Y_i \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} X_{ik} Y_i \end{bmatrix}$$

hence,

$$(2.45)\qquad \begin{aligned} b_0 &= c_{00} \sum_{i=1}^{n} Y_i + c_{01} \sum_{i=1}^{n} X_{i1} Y_i + c_{02} \sum_{i=1}^{n} X_{i2} Y_i + \cdots + c_{0k} \sum_{i=1}^{n} X_{ik} Y_i \\ b_1 &= c_{10} \sum_{i=1}^{n} Y_i + c_{11} \sum_{i=1}^{n} X_{i1} Y_i + c_{12} \sum_{i=1}^{n} X_{i2} Y_i + \cdots + c_{1k} \sum_{i=1}^{n} X_{ik} Y_i \\ &\cdots \\ b_k &= c_{k0} \sum_{i=1}^{n} Y_i + c_{k1} \sum_{i=1}^{n} X_{i1} Y_i + c_{k2} \sum_{i=1}^{n} X_{i2} Y_i + \cdots + c_{kk} \sum_{i=1}^{n} X_{ik} Y_i \end{aligned}$$

Some Measures of Variation

The variation about the origin in the endogenous variable is represented by the second moment about the origin, $Y'Y = \sum_{i=1}^{n} Y_i^2$. The amount of variation about the origin in Y accounted for by the (k+1) exogenous variables is $b'X'Y$ which in terms of the estimators (2.44) and cross

moments in (2.35) and 2.36) is:

$$(2.46)\quad b'X'Y' = b_0 \sum_{i=1}^{n} Y_i + b_1 \sum_{i=1}^{n} X_{i1}Y_i + b_2 \sum_{i=1}^{n} X_{i2}Y_i + \cdots + b_k \sum_{i=1}^{n} X_{ik}Y_i$$

Generally the variation of concern in the endogenous variable is not the variation about the origin but the total variation about the mean. In terms of moments this is the second moment about the mean. The total variation about the mean of the endogenous variable is frequently referred to as the total sums of squares and is the variation about the origin corrected for the amount attributable to the meal level.

$$(2.47)\quad y'y = \sum_{i=1}^{n} Y_i^2 = Y'Y - n\bar{Y}^2 = \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

The amount of variation in Y, measured about the mean of Y accounted for by the k exogenous variables is called the sum of squares regression and is as follows:

$$(2.48)\quad SSR = b'X'Y - n\bar{Y}^2$$

$$= b_0 \sum_{i=1}^{n} Y_i + b_1 \sum_{i=1}^{n} X_{i1}Y_i + b_2 \sum_{i=1}^{n} X_{i2}Y_i + \cdots + b_k \sum_{i=1}^{n} X_{ik}Y_i - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

This represents the amount of variation in Y that is "explained" by the variables included in the regression equation. As SSR approaches y'y the specified relation is basically doing a good job in explaining the variability in the endogenous variable and we have a good fitting equation.

It might be well to point out that if the moments are adjusted for means such as those (2.42) the sum of squares regression is:

$$(2.49) \quad SSR = b_1 \sum_{i=1}^{n} x_{i1}y_i + b_2 \sum_{i=1}^{n} x_{i2}y_i + \cdots + b_k \sum_{i=1}^{n} x_{ik}y_i$$

$$= [b_1, b_2, \cdots b_k]\,[x'y]$$

The $b_1, b_2, \cdots, b_k$ are the same as those given in (2.45) disregarding the estimator of the intercept.

The amount of variation in Y which is not accounted for by the exogenous variables (including the dummy for the mean) in the relation is the sum of squares error or residual. It is an indicator of the lack of fit of the regression equation. The sum of squares residual denoted as SSE is:

$$(2.50) \quad SSE = e'e = Y'Y - b'X'Y = y'y - SSR$$

It follows that the total variation is the sum of two components, $y'y = SSR + SSE$.

The least squares estimator of variance may be obtained from (2.50) by simply dividing the sum of squared residuals by the appropriate degrees of freedom, $n - k - 1$. Hence,

$$(2.51) \quad s^2 = \frac{e'e}{n - k - 1} = \frac{SSE}{n - k - 1}$$

One of the more common measures of goodness of fit is the R^2 or coefficient of determination. R^2 indicates the proportion of variation in the endogenous variable, where variation is measured in squared deviations about the mean, accounted for by the exogenous variables in the regression relation.

$$(2.52) \quad R^2 = \frac{b'X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2} = \frac{SRR}{y'y}$$

R^2 may take on values between 0 and one. An R^2 value near zero indicates that the exogenous variables in the regression equation are not very useful in terms of accounting for the variation in the endogenous variable. $R^2 = 1$ indicates a perfect fit which from an applied viewpoint is not only an uninteresting case but a candidate for suspicion. Perfect fits usually provide little useful information and often arise only because of inadequate degrees of freedom. High R^2 (but not equal to one) in cases where inadequate degrees of freedom are not a problem indicate a good fitting equation in which the exogenous variables account for a large proportion of the variation in the endogenous variable.

The coefficient of non-determination follows directly from the coefficient of determination. The coefficient of non-determination is

$$(2.53) \qquad 1 - R^2 = \frac{e'e}{Y'Y - n\bar{Y}^2} = \frac{e'e}{y'y} = \frac{SSE}{y'y}$$

It measures the proportion of the variation about the mean in the endogenous which is not explained or accounted for by the exogenous variables in the equation.

The coefficients of determination as expressed in (2.52) and non-determination given in (2.53) are not adjusted for differences in degrees of freedom associated with the numerator and denominator components. Basically sample size was used for both which cancelled out and left the resulting coefficients unaffected. However, the multiple correlation coefficient squared or coefficient of determination and the coefficient of non-determination may be defined in terms of degrees of freedom rather than sample size. Such coefficients are referred to as adjusted for degrees of freedom or simply adjusted. If the sample size

is very large and the difference between this and degrees of freedom is small (i.e., relative to the number of observations the number of parameters being estimated is small) the adjustment for degrees of freedom will be small. On the other hand, if the sample size is rather small and relative to the number of observations the number of parameters being estimated is large then the adjustment for degrees of freedom may be quite substantial. Denoting the adjusted coefficients of determination with a bar we have:

$$(2.54) \qquad 1 - \bar{R}^2 = \frac{\frac{e'e}{n-k-1}}{\frac{Y'Y - n\bar{Y}^2}{n-1}} = \frac{n-1}{n-k-1}\ (1 - R^2)$$

and solving for $\bar{R}^2$ yields

$$(2.55) \qquad \bar{R}^2 = 1 - \frac{n-1}{n-k-1}\ (1 - R^2)$$

The adjusted coefficient of determination is interpreted in a similar manner as the unadjusted coefficients. It is a measure of the proportion of variation in the endogenous variable accounted for by the exogenous variables included in the model after adjustments for degrees of freedom have been made.

Now let us consider the covariance matrix given in (2.25). $S(b) = s^2(X'X)^{-1}$ and letting $(X'X)^{-1}$ be C as given in (2.43) we have

$$(2.56) \qquad S(b) = s^2 \begin{bmatrix} c_{00} & c_{01} & c_{02} & - & - & - & - & c_{0k} \\ c_{10} & c_{11} & c_{12} & - & - & - & - & c_{1k} \\ c_{20} & c_{21} & c_{22} & - & - & - & - & c_{1k} \\ \cdot & \cdot & \cdot & & & & & \cdot \\ \cdot & \cdot & \cdot & & & & & \cdot \\ \cdot & \cdot & \cdot & & & & & \cdot \\ c_{k0} & c_{k1} & c_{k2} & & & & & c_{kk} \end{bmatrix}$$

The estimated variances of the estimators of regression parameters are the product of s^2 and the respective diagonal elements of C.[10/] For example the estimated intercept is $s^2_{b0} = c_{00}s^2$, of b_1 is $s^2_{b_1} = c_{11}s^2$, of b_2 is $s^2_{b_2} = c_{22}s^2_2$, etc. The covariances between pairs of estimated coefficients are given in the product of s^2 and the respective off diagonal elements. For example, the covariance of b_1 and b_2 is $s_{b_1b_2} = c_{12}s^2 = c_{21}s^2$ since $C = (X'X)^{-1}$ is symmetric. The covariance of any b_i and b_j is $s_{b_kb_j} = c_{ij}s^2 = c_{ji}s^2$.

This is a k x k matrix with the estimated variances on the major diagonal and covariances off-diagonal. However, the estimated variance of b_0 and the estimated covariances of b_0 with other estimated coefficients are not available from (2.57). (2.57) is the submatrix of (2.56) when the first row and first column of (2.56) are deleted.[11/]

Estimation from a Correlation Coefficient Viewpoint and Standardized Regression

Some computational procedures for obtaining least squares estimators involve transforming the moments about the means to correlation coefficients and perfoming the matrix inversion and other computations in terms

---

[10/] If the normal equations were formulated in terms of moments adjusted for means as expressed in (2.42) the inverse to be considered is $(x'x)-1$ where x represents the matrix of exogenous variables expressed as deviations from their respective means. The covariance matrix of the estimators is:

$$(2.57) \qquad S^*(b) = s^2 (x'x)^{-1}$$

[11/] Arthur S. Goldberger proves this on pages 182-184 of Econometric Theory, 1964, as does Henri Theil in Chapter 1 of Principles of Econometrics, 1971. Franklin Graybill proves the simple regression case on page 118 of An Introduction to Linear Statistical Models, Vol. 1, 1961.

of the correlation coefficients. One advantage of using the correlation coefficients is that the matrix for which an inverse has to be obtained is well conditioned in that all of its elements take on values between -1 and +1 since these are the bounds of product-moment correlations. This section will not be as detailed as the previous section since its purpose is to show the relationship between the moment approach and correlation coefficient approach to obtaining least squares estimators. The estimators which are first obtained by using the correlation coefficient approach are referred to as standard partial regression coefficients. The standard estimators may be very easily and directly transformed to the ordinary least squares estimators.

The product-moment correlation:

$$(2.58)\quad r = \frac{\dfrac{\sum_{i=1}^{n} (x_{i1} - \bar{X}_1)\ (X_{i2} - \bar{X}_2)}{n-1}}{\sqrt{\dfrac{\sum_{i=1}^{n} (X_{i1} - \bar{X}_1)^2}{n-1}} \cdot \sqrt{\dfrac{\sum_{i=1}^{n} (X_{i2} - \bar{X}_2)^2}{n-1}}} = \frac{\sum_{i=1}^{n} x_{i1}x_{i2}}{\sqrt{\sum_{i=1}^{n} x_{i1}^2} \cdot \sqrt{\sum_{i=1}^{n} x_{i2}^2}}$$

The correlation coefficient is a measure of how two variables vary together either in the same direction or opposite directions. Specifically, the correlation coefficient is a measure of the degree of linear association between two variables. It is a dimensionless quantity independent of any units of measurement. The extreme values of r, r=1 and r=-1, represent perfect positive and negative linear association,

respectively, r=0 indicates there is no linear association between the two variables. A linear correlation of zero does not preclude some type of non-linear relationship between the two variables.

Now let us consider the normal equations of the least squares procedure when expressed in terms of moments about the means as they were presented in (2.42). Unless specified otherwise, all summations are over i = 1, 2, . . ., n and hence subscripts are dropped. These normal equations are:

$$
\begin{aligned}
& b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 + - - - + b_k \Sigma x_1 x_k = \Sigma x_1 y \\
& b_1 \Sigma x_2 x_1 + b_2 \Sigma x_2^2 + - - - + b_k \Sigma x_2 x_k = \Sigma x_2 y \\
& \qquad . \qquad\qquad . \qquad\qquad . \qquad . \\
& \qquad . \qquad\qquad . \qquad\qquad . \qquad . \\
& \qquad . \qquad\qquad . \qquad\qquad . \qquad . \\
& b_1 \Sigma x_k x_1 + b_2 \Sigma x_k x_2 + - - - b_k \Sigma x_k^2 = \Sigma x_k y
\end{aligned}
\tag{2.59}
$$

Now let us divide both sides of the first equation of (2.59) by $\sqrt{\Sigma x_1^2} \cdot \sqrt{\Sigma y^2}$, both sides of the second equation by $\sqrt{\Sigma x_2^2} \cdot \sqrt{\Sigma y^2}$ etc. down to the kth equation which is divided by $\sqrt{\Sigma x_k^2} \sqrt{\Sigma y^2}$. The resulting set of equations is

$$
\begin{aligned}
& b_1 \frac{\sqrt{\Sigma x_1^2}}{\sqrt{\Sigma y^2}} + b_2 \frac{\Sigma x_1 x_2}{\sqrt{\Sigma x_1^2}\sqrt{\Sigma y^2}} + - - - - + b_k \frac{\Sigma x_1 x_k}{\sqrt{\Sigma x_1^2}\sqrt{\Sigma y^2}} = \frac{\Sigma x_1 y}{\sqrt{\Sigma x_1^2}\sqrt{\Sigma y^2}} \\
& b_1 \frac{\Sigma x_2 x_1}{\sqrt{\Sigma x_2^2}\sqrt{\Sigma y^2}} + b_2 \frac{\sqrt{\Sigma x_2^2}}{\sqrt{\Sigma y^2}} + - - - - + b_k \frac{\Sigma x_2 x_k}{\sqrt{\Sigma x_2^2}\sqrt{\Sigma y^2}} = \frac{\Sigma x_2 y}{\sqrt{\Sigma x_2^2}\sqrt{\Sigma y^2}} \\
& \qquad . \\
& \qquad . \\
& \qquad . \\
& b_1 \frac{\Sigma x_k x_1}{\sqrt{\Sigma x_k^2}\sqrt{\Sigma y^2}} + b_2 \frac{\Sigma x_k x_2}{\sqrt{\Sigma x_k^2}\sqrt{\Sigma y^2}} + - - - - + b_k \frac{\sqrt{\Sigma x_k^2}}{\sqrt{\Sigma y^2}} = \frac{\Sigma x_k y}{\sqrt{\Sigma x_k^2}\sqrt{\Sigma y^2}}
\end{aligned}
\tag{2.60}
$$

Let these estimators be denoted with an asterisk.

The relationship between these and the least squares estimators is:

$$(2.61) \quad b_j^* = b_j \frac{\sqrt{\Sigma x_j^2}}{\sqrt{\Sigma y^2}} = b_j \frac{\sqrt{\Sigma x_j^2}}{\sqrt{\Sigma y^2}} \cdot \frac{\sqrt{n-1}}{\sqrt{n-1}}$$

where b_j is the least squares estimator of the jth parameter and s_j and s_y are the standard errors of X_j and Y, respectively. It follows that

$$(2.62) \quad b_j = b_j^* \frac{\sqrt{\Sigma y^2}}{\sqrt{\Sigma x_j^2}} = b_j^* \frac{s_y}{s_j}$$

The b^*'s are called standard partial regression coefficients. They are independent of the original units of measurement and comparisons of any two indicates the relative importance of the two variables in terms of predicting values of the endogenous variable and accounting for variation in the endogenous variable.[12/] The notion of relative importance is strictly from a statistical viewpoint and in a mechanical sense.

Let us now rewrite the equations presented in (2.60) in terms of the standard estimators. This transformed system of equations is:

12/ For more details see Robert G. D. Steel and James Torrie, Principles and Procedures of Statistics, 1960, pages 284-85.

$$b_1^* + b_2^* \frac{\Sigma x_1 x_2}{\sqrt{\Sigma x_1^2}\sqrt{\Sigma x_2^2}} + - - - - - + b_k^* \frac{\Sigma x_1 x_k}{\sqrt{\Sigma x_1^2}\sqrt{\Sigma x_k^2}} = \frac{\Sigma x_1 y}{\sqrt{\Sigma x_1^2}\sqrt{\Sigma y^2}}$$

$$b_1^* \frac{\Sigma x_2 x_1}{\sqrt{\Sigma x_2^2}\sqrt{\Sigma x_1^2}} + b_2^* + - - - - - + b_k^* \frac{\Sigma x_2 x_k}{\sqrt{\Sigma x_2^2}\sqrt{\Sigma x_k^2}} = \frac{\Sigma x_2 y}{\sqrt{\Sigma x_2^2}\sqrt{\Sigma y^2}}$$

.

.

.

(2.63)

$$b_1^* \frac{\Sigma x_k x_1}{\sqrt{\Sigma x_k^2}\sqrt{\Sigma x_1^2}} + b_2^* \frac{\Sigma x_k x_2}{\sqrt{\Sigma x_k^2}\sqrt{\Sigma x_2^2}} + - - - - - + b_k^* = \frac{\Sigma x_k y}{\sqrt{\Sigma x_k^2}\sqrt{\Sigma y^2}}$$

But

$$\frac{\Sigma x_k x_1}{\sqrt{\Sigma x_1^2}\sqrt{\Sigma x_j^2}} = r_{ij}$$ which is the correlation coefficient

between the ith and jth variables and hence (2.63) may be rewritten in terms of correlation coefficients.

$$b_1^* + b_2^* r_{12} + - - - - - + b_k^* r_{1k} = r_{1y}$$

$$b_1^* r_{21} + b_2^* + - - - - - + b_k^* r_{2k} = r_{2y}$$

.

.

.

$$b_1^* r_{k1} + b_2^* r_{k2} + - - - - - + b_k^* = r_{ky}$$

which when written in terms of matrices is:

$$(2.64)\quad \begin{bmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & \cdots & r_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_k^* \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \\ \vdots \\ r_{ky} \end{bmatrix}$$

and in general notation

$$(2.65)\quad [r_{ij}]\, b^* = r_{jy}$$

where $[r_{ij}]$ is a k x k matrix of correlation coefficients of pairs of exogenous variables.

b^* is a k x 1 vector of standardized regression coefficients.

r_{jy} is a k x 1 vector of correlation coefficients between exogenous and endogenous variables.

The standard estimators are obtained by solving for b^* as follows:

$$(2.66)\quad b^* = [r_{ij}]^{-1} r_{jy}$$

Let

$$C^* = [r_{ij}]^{-1} = \begin{bmatrix} c_{11}^* & c_{12}^* & \cdots & c_{1k}^* \\ c_{21}^* & c_{22}^* & \cdots & c_{2k}^* \\ \vdots & \vdots & & \\ c_{k1}^* & c_{k2}^* & \cdots & c_{kk}^* \end{bmatrix}$$

Then

$$b^* = C^* r_{jy} \tag{2.68}$$

and

$$\begin{aligned} b_1^* &= c_{11}^* r_{1y} + c_{12}^* r_{2y} + - - - - + c_{1k}^* r_{ky} \\ b_2^* &= c_{21}^* r_{1y} + c_{22}^* r_{2y} + - - - - + c_{2k}^* r_{ky} \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \\ b_k^* &= c_{k1}^* r_{1y} + c_{k2}^* r_{2y} + - - - - + c_{kk}^* r_{ky} \end{aligned} \tag{2.69}$$

The estimators presented in (2.66), (2.68) and (2.69) are estimators of the coefficients of a standardized regression equation, a transformed version of the conventional one we were considering earlier, which may be expressed as:

$$y^* = b_1^* x_1^* + b_2^* x_2^* + - - - - + b_k^* x_k^* + e^* \tag{2.70}$$

where

$$y^* = \frac{(Y - \bar{Y})}{s_y} \quad \text{and} \quad Y = y^* s_y + \bar{Y} \tag{2.71}$$

$$x_j^* = \frac{(X_j - \bar{X}_j)}{s_j} \quad \text{and} \quad X_j = x_j^* s_j + \bar{X}_j \tag{2.72}$$

e^* is the residual

b_j^*'s are the standard least squares estimators.

Usually the standard regression equation given in (2.70) is not presented and when it is, it is usually expressed in prediction equation form. In prediction form, the standardized endogenous variable is replaced by standardized predicted values of the endogenous variables as follows:

$$\hat{y}^* = (y^* - e^*) = b_1^* x_1^* + b_2^* x_2^* + - - - + b_k^* x_k^* \tag{2.73}$$

where

$$(2.74) \qquad \hat{y}^* = \frac{(\hat{Y} - \bar{Y})}{s_y} \quad \text{and} \quad \hat{Y} = \hat{y}^* s_y + \bar{Y}$$

Other statistical quantities such as sum of squares due to regression, sum of squared residuals, coefficient of determination, etc. may be computed in the standardized framework. However, standardized quantities can be transformed to corresponding non-standardized quantities quite easily. The transformation for parameter estimators is given by (2.62) and for the endogenous, exogenous and predicted endogenous variables by (2.71), (2.72) and (2.74), respectively. The inverse of the standardized form C^* may be transformed to the inverse of the nonstandardized form involving moments about the means by the following;

$$(2.75) \qquad c^*_{ij} = c_{ij} \frac{\Sigma x_i x_j}{r_{ij}} = c_{ij} \Sigma x_i^2 \Sigma x_j^2$$

and

$$(2.76) \qquad c_{ij} = c^*_{ij} \frac{r_{ij}}{\Sigma x_i x_j} = \frac{c_{ij}}{\Sigma x_i^2 \quad \Sigma x_j^2}$$

where

c^*_{ij} is the element in the ith row and jth column of C^*.

c_{ij} is the element in the ith row and jth column of $(x'x)^{-1}$.

Note that if $\Sigma x_i x_j = 0$ then $r_{ij} = 0$ and the first expressions to the right of the equal signs in (2.75) and (2.76) cannot be used for computational purposes.

After the inverse of the matrix of correlation coefficients and estimates of the standardized regression coefficients have been obtained, the standardized quantities may be transformed into non-standardized forms for the purpose of computing other quantities. Testing of hypotheses and other analyses may be performed in the non-standardized form. After obtaining the inverse and standardized regression coefficients, there is no apparent advantage for continuing to use the standardized framework.

Interval Estimation

Statistical inference is concerned with two types of problems - estimation and testing of hypothese. Estimation entails two types of estimators, point and interval. The least squares estimators, b, which we had discussed earlier are point estimators. Point estimators yield a single numerical estimate of each parameter being estimated. There is no specific probability of occurrence associated with the single valued numerical estimate.

In estimation, the estimates we obtain usually do not have exactly the same value of the parameter being estimated. It makes sense to give the estimates in terms of an interval and these intervals are called confidence intervals. The interval is given in terms of a specific probability level. Confidence interval gives the shortest range or interval which at a specified probability level is expected to contain or include the parameter for which the interval is being established.

Three types of interval estimators related to least squares multiple regression will be considered. First, the interval estimator (confidence limits) for a single regression parameter will be presented. Next, the interval estimator for the difference between a pair of regression parameters will be given. Finally, the interval estimator for an average predicted value of the endogenous variable and for an individual value of the endogenous variable will be presented and discussed.

Confidence Limits for Individual Regression Parameters

In setting conficence limits upon a specific β_j we wish to establish an interval such as

$$(2.77) \quad P(L_1 < \beta_j < L_2) = 1 - \alpha$$

where the probability of the interval (L_1, L_2) containing the population parameter β_j is equal to $(1 - \alpha)$.[13/] $(1 - \alpha)$ is the level of confidence which is one minus the level of significance where α is the level of significance and the probability of committing a Type I error. L_1 and L_2 are the limits which are to be established.

In establishing a conficence interval for any parameter values from a known probability distribution are utilized. The particular theoretical values are determined by the selected probability level, $(1 - \alpha)$. For example, .95 confidence limits ($\alpha = .05$) would mean that the theoretical values of some known probability distribution would be chosen or greater than the larger value chosen is .05. The particular known probability distribution chosen to establish confidence limits for a particular type of parameter is determined by the sampling distribution of some function of the parameter for which the interval estimate is being made. In the case of β_j the function we are concerned with is:

$$\frac{b_j - \beta_j}{s_{b_j}} \tag{2.78}$$

which is distributed as a t. In general, the difference between an estimator and the parameter over the standard error of the estimator is distributed as a t. The general probability expression of the confidence

---

[13/] This is an open interval and hence does not include the boundary points. A closed interval could be chosen and the boundary points would be included. The probability expression would contain the weak inequality and (2.77) would be:

$$P(L_1 \leq \beta_j \leq L_2) = 1 - \alpha$$

interval is:

$$(2.79) \quad P\left[-t_{\alpha/2} < \frac{b_j - \beta_j}{s_{b_j}} < t_{\alpha/2}\right] = 1 - \alpha$$

$$P\left[-t_{\alpha/2} s_{b_j} < b_j - \beta_j < t_{\alpha/2}\, s_{b_j}\right] = 1 - \alpha$$

$$P\left[-b_j - t_{\alpha/2}\, s_{b_j} < -\beta_j < -b_j + t_{\alpha/2}\, s_{b_j}\right] = 1 - \alpha$$

$$P\left[b_j + t_{\alpha/2}\, s_{b_j} > \beta_j > b_j - t_{\alpha/2}\, s_{b_j}\right] = 1 - \alpha$$

$$P\left[b_j - t_{\alpha/2}\, s_{b_j} < \beta_j < b_j + t_{\alpha/2}\, s_{b_j}\right] = 1 - \alpha$$

Hence:

$$(2.80) \quad L_1 = b_j - t_{\alpha/2}\, s_{b_j}$$

$$L_2 = b_j + t_{\alpha/2}\, s_{b_j}$$

where L_1 is the lower limit

L_2 is the upper limit

b_j is the least squares estimator of β_j.

$t_{\alpha/2}$ is the tabled value of t for $\alpha/2$ probability level and the same degrees of freedom as s^2.

$s_{b_j} = \sqrt{s^2 c_{jj}}$ is the standard error of b_j.

Note that the two-tailed t was used. Both a lower and upper limit are being established and part of the α probability is in each of the tails, hence the $\alpha/2$. The two theorectical values are used except in cases where one limit is known a priori and only one limit needs to be established on a probabilistic basis.

An appropriate interpretation of a confidence interval is that the probability is $(1 - \alpha)$ that the interval established by (2.80) contains the parameter β_j. Alternatively, one is 100 $(1 - \alpha)$ percent confident that the interval established by (2.80) contains the true parameter. These statements apply to confidence intervals in general.

There are two things worth noting about the confidence interval statement. First, the statement is made in such a way that the parameter is a constant and does not vary but the confidence limits vary in repeated sampling. Second, from a probability viewpoint the statement is valid only before numerical values are established for the confidence limits. Once the numerical values are established, the probability that the parameter is contained within the limits is either one or zero.

Confidence Limits for the Difference Between Two Regression Parameters

Following the same procedure as the single regression parameter we can quickly established the confidence interval for the difference between two regression parameters. The quantity we are concerned with is

$$(2.81) \quad \frac{\left[(b_j - \beta_j) - (b_r - \beta_r)\right]}{s_{(b_j - b_r)}} = \frac{\left[(b_j - b_r) - (\beta_j - \beta_r)\right]}{s_{(b_j - b_r)}}$$

The confidence interval is

$$(2.82) \quad P\left[(b_j - b_r) - t_{\alpha/2}\, s_{(b_j - b_r)} < (\beta_j - \beta_r) < (b_j - b_r) + t_{\alpha/2} s_{(b_j - b_r)}\right] = 1 - \alpha$$

and

$$(2.83)\quad L_1 = (b_j - b_r) - t_{\alpha/2}\, s_{(b_j - b_r)}$$

$$L_2 = (b_j - b_r) + t_{\alpha/2}\, s_{(b_j - b_r)}$$

where:

L_1 is the lower limit

L_2 is the upper limit

$(b_j - b_r)$ is the difference between the jth and the rth estimators

$t_{\alpha/2}$ is the tabled t value for $\alpha/2$ probability level and the same degrees of freedom as s^2

$s_{(b_j - b_r)} = \sqrt{s^2 (c_{jj} + c_{rr} - 2c_{jr})}$ is the standard error of the difference between the jth and rth estimators.

Confidence Limits for Predictions

Sometimes linear multiple regression equations are estimated primarily for the purpose of predicting values of the endogenous variable. In many instances an interval estimate may be more enlightening and useful than point estimates of the endogenous variable. For example, it might be more useful to know that the probability is .95, that the price of corn will be between \$2.10 and \$2.20 per bushel for a set of exogenous variables than to only know that for that set of exogenous variables the prediction is \$2.15 per bushel. The latter prediction provides no information about the chances of the predicted value occurring. However, if one is going to establish confidence limits for the price of corn per bushel one has to decide whether the limits will be set on the average price per bushel

during a specified time or on the price of any individual bushel of corn which might be sold during some specified time period. The two types of confidence limits for predicted values are:

(1) Confidence limits for the average value of the endogenous variable which might occur for a given set of values of the enogenous variables.

(2) The confidence limits for any single predicted value of the endogenous variable for a given set of values of the exogenous variables.

Let us consider the linear multiple regression model in its prediction form i.e., $\hat{Y} = Y - e$ then

$$(2.84)\quad \hat{Y} = b_0 + b_1X_1 + b_2X_3 + - - - + b_kX_k$$

Now let us obtain a predicted value of the endogenous variable for a specific set of X's which will be denoted with a super +. By a specific set of X's is meant a specific value of each X from X_1 through X_k. The predicted value of Y for the specific set of X's may be expressed in equation form as follows:

$$(2.85)\quad \hat{Y}^+ = b_0 + b_1X_1^+ + b_2X_2^+ + b_3X_3^+ + - - - + b_kX_k^+$$

$$= \bar{Y} + b_1(X_1^+ - \bar{X}_1) + b_2(X_2^+ - \bar{X}_2) + b_3(X_3^+ - \bar{X}_3) + - - - b_k(X_k^+ - \bar{X}_k)$$

$$= \bar{Y} + b_1x_1^+ + b_2x_2^+ + b_3x_3^+ + - - - + b_kx_k^+$$

where:

$\hat{Y}^+$ is the predicted value of the endogenous variable

X_j^+ is the chosen value of the jth exogenous variable

x_j^+ is the deviation of the chosen value of the jth exogenous variable from the mean of the jth exogenous variable

b_j is the jth least squares estimator.

The estimated variance of an average predicted Y is

$$(2.86) \quad s^2_{\hat{Y}^+} = s^2 \left[\frac{1}{n} + \sum_{i=1}^{k} \sum_{j=1}^{k} c_{ij} x_i^+ x_j^+\right]$$

where:

s^2 is the residual variance, i.e., $\frac{e'e}{n-k-1}$

c_{ij} is the element in the ith row and jth column of $(X'X)^{-1}$

n is the number of observations, i.e., sample size

For a three exogenous variable example, the estimated variance of an average predicted Y is

$$(2.87) \quad s^2_{\hat{Y}^+} = s^2 \left[\frac{1}{n} + x_1^+ (c_{11}x_1^+ + c_{12}x_2^+ + c_{13}x_3^+) + x_2^+ (c_{21}x_1^+ + c_{22}x_2^+ + c_{23}x_3^+) + x_3^+ (c_{31}x_1^+ + c_{32}x_2^+ + c_{33}x_3^+)\right]$$

Note there are two components of variation in $s^2_{\hat{Y}^+}$. The first is the variance of the mean which evolves from $\frac{s^2}{n}$. The second is the sampling variability and covariability of the estimated regression coefficients weighted by the squares and cross products of deviation of the chosen values of the exogenous variables from their respective means.[14/]

[14/] This point is more easily seen in the simple regression case when there is only one exogenous variable. Then

$$s^2_{\hat{Y}^+} = s^2 \left[\frac{1}{n} + \frac{(x^+)^2}{\Sigma x^2}\right] = \left[\frac{s^2}{n} + s^2_b (x^+)^2\right]$$

This latter component becomes a more and more dominant factor as the chosen values of the exogenous variables deviate further and further away from their respective mean values and the result is a larger and larger estimated variance of the predicted average value of the endogenous variable.

Similar to the confidence limits for the regression coefficients, the quantity of interest is $\frac{Y^+ - E(Y^+)}{s_{\hat{Y}^+}}$ which is distributed as a t.

The $(1 - \alpha)$ confidence limits are:

$$(2.88) \quad P\left(\hat{Y}^+ - t_{\alpha/2}\ s_{\hat{Y}^+} < E(Y^+) < \hat{Y}^+ + t_{\alpha/2}\ s_{\hat{Y}^+}\right) = 1 - \alpha$$

and

$$(2.89) \quad L_1 = \hat{Y}^+ - t_{\alpha/2}\ s_{\hat{Y}^+}$$

$$L_2 = \hat{Y}^+ + t_{\alpha/2}\ s_{\hat{Y}^+}$$

where:

L_1 is the lower limit

L_2 is the upper limit

$t_{\alpha/2}$ is the tabled t value for the $\alpha/2$ probability level and the same degrees of freedom as s^2.

$s_{\hat{Y}^+} = \sqrt{s^2_{\hat{Y}^+}}$ as defined in (2.86) and illustrated in (2.87)

The confidence limits for any single predicted value of Y for chosen values of all of the exogenous variables are determined in a similar manner but the estimated variance of the single predicted value of Y includes an additional component of variation. The third component of variation is the random variation of individual items as measured by the residual variance.

The residual variance is a measure of the random variation among individual items even though treated alike and subjected to the same stimuli.

The variance of a single predicted value of Y is

$$(2.90) \quad s_{Y^+}^2 = s^2 \left[1 + \frac{1}{n} + \sum_{i=1}^{k} \sum_{j=1}^{k} c_{ij} x_i^+ x_j^+\right]$$

All terms have already been defined. The three variable illustrative example would be the same as (2.87) except for adding a one inside the brackets.

The $(1 - \alpha)$ confidence limits are:

$$(2.91) \quad P\left[\hat{Y}^+ - t_{\alpha/2}\, s_{Y^+} < Y^+ < \hat{Y}^+ + t_{\alpha/2}\, s_{Y^+}\right] = 1 - \alpha$$

and

$$(2.92) \quad L_1 = \hat{Y}^+ - t_{\alpha/2}\, s_{Y^+}$$

$$L_2 = \hat{Y}^+ + t_{\alpha/2}\, s_{Y^+}$$

where:

L_1 is the lower limit

L_2 is the upper limit

$t_{\alpha/2}$ is the tabled value of t for the $\alpha/2$ probability level and the same degrees of freedom as s^2

$s_{Y^+} = \sqrt{s^2_{Y^+}}$ where s_{Y}^2 is defined in (2.90)

The confidence limits for the single predicted Y establish an interval such that the probability is $(1 - \alpha)$ that the true value of Y for the chosen levels of exogenous variables is contained in the interval. For the average predicted Y the interpretation is that the probability is $(1 - \alpha)$ that the interval contains the true <u>expected value</u> of the endogenous variable for the chosen values of the exogenous variables.

Some Numerical Examples

To complement the theoretical and computational frameworks presented thus far, attention is focused on two types of numerical examples. These numerical examples illustrate some aspects of estimation of the least squares linear multiple regression model. The first of the examples emphasizes estimation of regression coefficients using correlation coefficients. The second example focuses on point and interval estimation of parameters of a multiplicative relationship which is linear in logarithms.

Estimation Using Correlation Coefficients

Let us consider a study of the demand for a particular commodity, based on hypothetical annual data for 21 years, which yields the following results.[15/]

Means	Standard Deviation	Correlation Coefficients
$\bar{X}$ 51.843	$s_X = 9.205$	$r_{XY} = -0.9158$
$\bar{Y}$ 8.313	$s_Y = 1.780$	$r_{YT} = -0.8696$
$\bar{T} = 0$	$s_T = 6.057$	$r_{XT} = 0.9304$

where

X = per capita consumption in pounds

Y = deflated price in cents per pound

T = time in years.

15/ This example is problem 5.1 in J. Johnston, Econometric Methods, Second Edition, 1972, p. 170.

Note that for $\bar{T} = 0$ time is coded negatively from the 11th year back toward the beginning of the series and positively from the 11th year forward.

Let us consider the problem of computing the coefficient of time in the estimated relationship between Y and X and T and testing hypotheses about whether the coefficient differs significantly from zero. Y will be assumed to be an endogenous variable while X and T will be considered exogenous. Hence, we are considering a demand relation in which the quantity is assumed to be predetermined and adjustments to achieve equilibrium occur in the price variable.

Since the partial information presented is in standard deviations and correlation coefficients, it is advantageous to perform the estimation using the correlation matrix. Inverting the correlation matrix instead of the moment matrix and using this matrix yields standardized partial regression coefficients. As we have indicated before the standardized partial regression coefficients can be readily transformed to the ordinary least squares coefficients. Alternatively, one could transform the standard deviations and correlation coefficients into moments and cross-moments about the means and formulate the normal equations in terms of moments about the means.

Our interest at this point is to perform the estimation in terms of correlation coefficients. The normal equations of interest are:

$$b_1^* + b_2^* r_{XT} = r_{XY}$$

(2.93)

$$b_1^* r_{XT} + b_2^* = r_{TY}$$

which in matrix form are:

$$(2.94)\qquad \begin{bmatrix} 1 & r_{XT} \\ r_{XT} & 1 \end{bmatrix} \cdot \begin{bmatrix} b_1^* \\ b_2^* \end{bmatrix} = \begin{bmatrix} r_{XY} \\ r_{TY} \end{bmatrix}$$

where b_1^* and b_2^* are estimated standardized partial regression coefficients.

The standardized estimators are:

$$(2.95)\qquad \begin{bmatrix} b_1^* \\ b_2^* \end{bmatrix} = \begin{bmatrix} 1 & r_{XT} \\ r_{XT} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{XY} \\ r_{TY} \end{bmatrix}$$

The determinant of the correlation coefficient matrix is

$$(2.96)\qquad \begin{bmatrix} 1 & r_{XT} \\ r_{XT} & 1 \end{bmatrix} = 1 - r_{XT}^2 = 1 - (.9304) = 1 - .8656 = \underline{.1344}$$

and

$$(2.97)\qquad \begin{bmatrix} 1 & r_{XT} \\ r_{XT} & 1 \end{bmatrix}^{-1} = \frac{1}{.1344} \begin{bmatrix} 1 & -r_{XT} \\ -r_{XT} & 1 \end{bmatrix}$$

$$= \frac{1}{.1344} \begin{bmatrix} 1 & -.9304 \\ -.9304 & 1 \end{bmatrix} = \begin{bmatrix} 7.4405 & -6.9226 \\ -6.9226 & 7.4405 \end{bmatrix}$$

Hence:

$$(2.98)\qquad \begin{bmatrix} b_1^* \\ b_2^* \end{bmatrix} = \begin{bmatrix} 7.4405 \\ -6.9226 \end{bmatrix} \begin{bmatrix} -6.9226 \\ 7.4405 \end{bmatrix} \begin{bmatrix} -.9158 \\ -.8696 \end{bmatrix} = \begin{bmatrix} -.7941 \\ -.1305 \end{bmatrix}$$

Now to obtain the ordinary least squares estimators we adjust the standardized partial regression coefficients by the ratio of the standard error of the endogenous variable over the standard error of the respective exogenous variable.

$$(2.99) \quad b_1 = b_1^* \frac{s_Y}{s_X} = -.7941 \frac{1.730}{9.205} = -.1536$$

$$b_2 = b_2^* \frac{s_Y}{s_T} = -.1305 \frac{1.730}{6.057} = -.0384$$

The estimated intercept is:

$$(2.100) \quad b_0 = \bar{Y} - b_1\bar{X} - b_2T_2 = 8.313 - (-.1536)\ (51.643) - (-.0384)\ (0)$$

$$b_0 = 8.313 + 7.963$$

$$b_0 = 16.276$$

The estimated regression equation is:

$$(2.101) \quad \hat{Y} = 16.275 - .1536X - .0384\ T$$

The cross moment about the means for X and Y is

$$(2.102) \quad \Sigma xy = r_{XY}\, s_Y\, s_X\, (n-1) = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} \sqrt{\frac{\Sigma y^2}{n-1}}\ \sqrt{\frac{\Sigma x^2}{n-1}}(n-1)$$

$$= \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} = \sqrt{\frac{\Sigma y^2}{n-1}}\ \sqrt{\frac{\Sigma x^2}{n-1}}(n-1)$$

Hence

$$(2.103) \quad \Sigma xy = (-.00158)\ (1.730)\ (9.205)\ (20) = \underline{-300.1058}$$

The cross moment about the means of Y and T is:

(2.104) $\Sigma yt = r_{XT}\, s_Y\, s_T\, (n-1) = (-.8696)\ (1.78)\ (6.057)\ (20) = -187.5112$

The second moment about the mean of the endogenous variable is:

(2.105) $\Sigma y^2 = (n-1)\, s_Y^2 = 20\ (1.78)^2 = 20\ (3.1684) = \underline{63.3680}$

The amount of variation in the endogenous variable accounted for by the regression is

(2.106) $SSR = b_1\, \Sigma xy + b_2\, \Sigma yt$

$= (-.1536)\ (-300.1058) + (-.0384)\ (-187.5112)$

$= 53.2967$

The amount of variation in the endogenous variable which is not accounted for by the regression is:

(2.107) $SSE = \Sigma y^2 = SSR = 63.3680 - 53.2967 = \underline{10.0713}$

The proportion of variation in price accounted for by the per capita quantity and time is

(2.108) $R^2 = \frac{53.2967}{63.3680} \doteq .841$

About 84% of the variation in price is accounted for by the per capita quantity and time together. Hence, in terms of the proportion of variation accounted for this is a fairly good fitting function.

The coefficient of determination adjusted for degrees of freedom is

(2.109) $R^2 = 1 - \frac{n-1}{n-k-1}\ (1-R^2) = 1 - \frac{20}{18}\ (1 - .841) = 1 - \frac{20}{18}\ (.159)$

$= 1 - .1767$

$R^2 = .8233$

After adjusting for degrees of freedom the proportion of variation in the endogenous variable accounted for by the two exogenous variables is still 82.3 percent.

An estimate of the residual variance may be obtained from the error sum of squares given in (2.107). The estimated residual variance is

$$(2.110) \qquad s^2 = \frac{SSE}{n-k-1} = \frac{10.0713}{18} = .5595$$

In order to obtain the estimated variances of the b coefficients the diagonal elements of the inverse of the matrix of moments about the means are needed. These are obtained by adjusting the elements of the standardized regression in the following way:

$$(2.111) \quad c_{11} = \frac{c^*_{11} \; r_{11}}{\Sigma x^2} = \frac{c^*_{11} \; r_{11}}{s_X^2 \; (n-1)} = \frac{c^*_{11} \; (1)}{\frac{\Sigma x^2}{n-1} \; (n-1)} = \frac{c^*_{11}}{\Sigma x^2}$$

$$= \frac{(7.4405) \; (1)}{(9.205)^2 \; (20)} = \frac{7.4405}{84.7320 \; (20)} = \frac{7.4405}{1694.64} = \underline{.00439}$$

$$(2.112) \quad c_{22} = \frac{(7.4405) \; (1)}{(6.057)^2 \; (20)} = \frac{7.4405}{36.6872 \; (20)} = \frac{7.4405}{733.744} = \underline{.0104}$$

The variances of the estimates are

$$(2.113) \qquad s^2_{b_1} = c_{11} \; s^2 = (.00439) \; (.5595) = \underline{.002456}$$

$$s^2_{b_2} = c_{22} \; s^2 = (.01014) \; (.5595) = \underline{.005673}$$

and the standard errors of the estimated coefficients are:

$$(2.114) \qquad s_{b_1} = \sqrt{c_{11}\, s^2} = \sqrt{.002456} = \underline{.0496}$$

$$s_{b_2} = \sqrt{c_{22}\, s^2} = \sqrt{.005673} = \underline{.0753}$$

Now we can proceed to testing hypotheses about the parameter of the time variable.[16/] In this case there is inadequate prior information to justify a one-tailed alternative hypothesis and hence a two tailed alternative will be used. The appropriate hypotheses are:

$$HO : \beta_2 = 0$$

$$HA \quad \beta_2 \neq 0$$

The appropriate test is a t-test

$$(2.115) \qquad t = \frac{b_2 - \beta_2}{s_{b_2}} = \frac{-.0384 - 0}{.0753} = \frac{-.0384}{.0753} = -.509$$

To be significant at the .05 level the computed would have to be less than

$$t_{(.025)\ (18)} = -2.101$$

Hence, time whatever it is measuring is not a significant variable in accounting for the variation in the price of this commodity.

Time could be measuring any number of variables. One could agree that time is a "proxy" variable representing a measure of changes in tastes and preferences for this particular commodity. Over time changes

---

[16/] Although the theory and procedures of hypothesis testing are discussed in the next chapter it is convenient and hopefully meaningful to present the use of the single t-test as the final part of this problem.

in tastes and preferences are believed to be gradual and usually in a given direction rather than an erratic back and forth changing and hence time may be a fairly reasonable measure of such changes. Also, income is not included in this relation and incomes tend to change steadily in recent times and one might argue that the time variable is a "proxy" for income and other variables which show a steady pattern of change over time.

Point and Interval Estimation of Coefficients of Multiplicative Functions

The following is a matrix of hypothetical variances and covariances of the three variables:

X_1 = log food consumption per capita

X_2 = log food price

X_3 = log disposable income per capita

The variance-covariance matrix based on a sample of 20 observations is:

(2.116)

	X_1	X_2	X_3
X_1	11.50	-3.12	26.99
X_2		29.16	30.80
X_3			133.00

On the assumption that the demand relationship may be adequately represented by a function of the form $Y_1 = \beta_0 Y_2^{\beta_1} Y_3^{\beta_2}$ (where $X_1 = \log Y_1$), the problem is to estimate the income elasticity of demand and compute a 95 percent confidence interval for this parameter.

The model whose parameters we are interested in estimating is:

(2.117) $X_1 = \beta_0 + \beta_1 X_2 + \beta_2 X_3 + \varepsilon$

where:

$X_1 = \log Y_1$

$X_2 = \log Y_2$

$X_3 = \log Y_3$

ε is a random disturbance in logarithmic terms

β_0, β_1, and β_2 are the regression parameters with β_0 being the intercept

In terms of logs (2.117) is:

(2.118) $\log Y_1 = \beta_0 + \beta_1 \log Y_2 + \beta_2 \log Y_3 + \varepsilon$

In logarithms the model is linear and estimation may be performed in terms of conventional least squares methods. The least squares estimators are elasticities. This may be shown by considering the multiplicative form of (2.118) and focusing on β_2, the coefficient of the per capita disposable income variable. An income elasticity of demand is defined as a relative change in the quantity consumed (demanded) over a relative change in income. Hence, in terms of partial derivatives the income elasticity of demand is:

(2.119) $$E_I = \frac{\frac{\partial Y_1}{Y_1}}{\frac{\partial Y_3}{Y_3}} = \frac{\partial Y_1}{\partial Y_3} \cdot \frac{Y_3}{Y_1}$$

But

(2.120) $$Y_1 = \beta_0 Y_2^{\beta_1} Y_3^{\beta_2} \varepsilon$$

and

$$(2.121) \qquad \frac{\partial Y_1}{\partial Y_3} = \beta_2 \beta_0 Y_2^{\beta_1} Y_3^{\beta_2 - 1} \epsilon$$

Now

$$(2.122) \qquad E_I = \frac{\partial Y_1}{\partial Y_3} \cdot \frac{Y_3}{Y_1} = \beta_2 \beta_0 Y_2^{\beta_1} Y_3^{\beta_2 - 1} \epsilon \frac{Y_3}{Y_1}$$

$$= \beta_2 \beta_0 Y_2^{\beta_1} Y_3^{\beta_2} \epsilon = \frac{\beta_2 Y_1}{Y_1} = \beta_2$$

Without loss of generalization the elasticity concept applies to each of the parameters of a multiplicative relation which is linear in logarithms.

Turning to the estimation of the coefficients the determinant of the variance-covariance matrix is:

$$(2.123) \qquad \begin{vmatrix} 29.16 & 30.80 \\ 30.80 & 133.00 \end{vmatrix} = 2929.64$$

and the inverse of the variance covariance matrix is:

$$(2.124) \qquad [x'x]^{-1} = \frac{1}{2929.64} \begin{bmatrix} 133.00 & -30.80 \\ -30.80 & 29.16 \end{bmatrix} = \begin{bmatrix} .04540 & -.01051 \\ -.01051 & .00995 \end{bmatrix}$$

The estimates of the regression coefficients are:

$$(2.125) \qquad \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} .04540 & .01051 \\ -.01051 & .00995 \end{bmatrix} \begin{bmatrix} -3.12 \\ 26.99 \end{bmatrix} = \begin{bmatrix} -.42531 \\ .30134 \end{bmatrix}$$

The sum of squares regression is:

$$(2.126) \quad SSR = b_1 \Sigma x_1 x_2 + b_2 \Sigma x_1 x_3$$

$$= .42531 \ (-3.12) + .30134 \ (26.99) = 9.46013$$

The sum of squared residuals is:

$$(2.127) \quad SSE = \Sigma x_1^2 - SSR = 11.59 - 9.46 = 2.13$$

and the estimated residual variance is:

$$(2.128) \quad s^2 = \frac{2.13}{17} = .1253$$

The variance of the stimated coefficient of the disposable income variable is:

$$(2.129) \quad s^2_{b_2} = c_{ss} s^2 = .00995 \ (.1253) = .00124674$$

and the standard error of b_2 is:

$$(2.130) \quad s_{b_2} = \sqrt{c_{22} s^2} = .035$$

The estimated price elasticity of demand for food is $b_1 = -.425$ and the estimated income elasticity of demand for food is $b_2 = .301$. The 95 percent confidence limits for the income elasticity are:

$$(2.131) \quad b_2 - t_{(.025)\ (17)}\, s_{b_2} < \beta_2 < b_2 + t_{(.025)\ (17)}\, s_{b_2}$$

$$.301 - (2.110) \ (.035) < \beta_2 < .301 + (2.110) \ (.035)$$

$$.301 - .074 < \beta < .301 + .074$$

$$.227 < \beta < .375$$

Hence the probability is .95 or we are 95 percent confident that the income elasticity for food is between .227 and .375. In similar manner confidence limits could be set on the price elasticity of demand. In general, this is the procedure for making interval estimates for the individual parameters of a linear multiple regression equation.

Chapter 3 - Testing Hypotheses About Linear Regression Parameters

Introduction

In the previous chapter the general linear multiple regression model was presented and some general estimation procedures were given. In this chapter attention is focused on various types of hypotheses about regression parameters that may be tested and the appropriate tests for testing such hypotheses. First, some general concepts concerning testing of hypotheses will be discussed. Then hypotheses and tests concerned with accounted for variation, variation associated with deletion of variables, single coefficients, and pairs of coefficients will be considered. Finally, testing linear combinations of regression coefficients and equality among coefficients in two or more relations are discussed.

Fundamentals of Hypothesis Testing

Statistical hypotheses specify the value of a parameter, values for a set of parameters, or the relationship among two or more parameters. In testing statistical hypotheses, the investigator is confronted with the possibility of making one of two types of errors. If the specificationsof the null hypotheses are true but called false, i.e., the null hypothesis is rejected when in fact it is true, a Type I error is committed. A second type of error which may be committed is the Type II[1/] and it occurs when a null hypothesis which is false is not

---

1/ Closely associated with the Type II error is the concept of the power of a test. The power of a test is one minus the probability of committing a Type II error. The power of a test indicates the ability of the test to detect the alternative hypothesis when the null hypothesis is false and the specific alternative under consideration is true. The tests which will be considered are some of the more powerful tests available for the types of tests in which we are interested.

rejected. A Type I error may occur only if the null hypothesis is rejected and a Type II may occur only if the null hypothesis is not rejected. Hence, for any specific testing situation, the investigator is subject to committing one or the other of these errors but never both at the same time. Of course, no error is committed if a false null hypothesis is rejected or a true null hypothesis is not rejected. From a practical viewpoint it is usually either not feasible or impossible to verify whether an error was committed in rejecting or not rejecting the null hypothesis, whichever the case may be.

In testing of hypotheses the probability of occurrence of a Type I error is commonly referred to as the level of significance. Usually an acceptable level of significance is chosen prior to performing the statistical tests. Alternatively, results of statistical tests are sometimes reported in terms of the probability level at which the results of the tests are significant. The choice of a level of significance depends on how crucial it is for the problem being considered to call a null hypothesis false when it really is true.

For a specified level of significance the size of the rejection region is influenced by whether a one or two tailed alternative hypothesis is considered. All other things equal a smaller value of testing statistic is required to be significant (in the rejection region) at a specified level when a one tailed alternative is used as compared to a two tailed alternative hypothesis. Alternatively, a specified value of a testing statistic will be significant at a higher level (smaller probability of occurrence of Type I error) for the one tailed alternative hypothesis than for the two tailed given that everything else is the same. With one tailed alternative hypotheses and corresponding appropriate tests, smaller differences may be judged in the significant class than with a two tailed alternative and correspondingly appropriate tests and hence

it is to the interest of the investigator to use a one tailed alternative hypothesis whenever it is appropriate. The choice of the alternative hypothesis is based on:

1. Theory of the discipline and information about the directional flow of causality of the phenomena being investigated.

2. Information about the particular phenomena based on other research and investigations and general knowledge about the particular characteristics being tested.

3. Area of interest or importance for the particular problem being investigated.

Some of the basic and fundamental statistical tests of the least squares multiple regression frameworkare presented in this section. Little attention is devoted to the derivation of the statistical tests. Emphasis is on the computational aspects of the statistical tests. All the tests considered involve the regression parameters, β. Although in some cases the Central Limit Theorem could be relied upon to provide information about sampling distribution of estimators so that appropriate statistical tests could be performed, we will directly employ assumption (A.6) in Chapter 2 which specifies that the ε's are normally distributed.

Tests Concerned With Accounted For Variation

The total variation of the endogenous variable measured about its mean may be separated into two parts - that accounted for by the exogenous variables in the regression equation and the amount not accounted for by the exogenous variables in the regression equation. Each of these amounts of variation is a sum of squares. Furthermore, the total variation is also a sum of squares - the sum of squared deviations about the mean. The sampling distribution of sums of squares is a chi-square distribution. The ratio of two independently distributed chi squares is a F ratio. Hence, any tests performed in the analysis of accounted for variation will entail some type of ratio of sums of squares; generally, sums of squares divided by degrees of freedom.

Frequently, the question one wants to answer after estimating a regression relationship is whether a significant amount of variation in the endogenous variable was accounted for by the regression. Another way of stating this is to determine whether some of the regression coefficients are not equal to zero. The null and alternative hypothesis may be specified as follows.

$$HO: \quad \beta^{\Delta} = 0 \text{ where } \beta^{\Delta} \text{ excludes } \beta_0$$

$$HA: \quad \text{Not all } \beta_j = 0 \text{ or } \beta^{\Delta} \neq 0 \quad j = 1,2,3,\text{---},k$$

The analysis of variance table which provides all the needed information is as follows:

(3.1)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Due to regression	k	$SSR = b'X'Y - n\bar{Y}^2$	$MSR = SSR/k$
Residual (from regression)	n-k-1	$SSE = e'e = Y'Y - b'X'Y$	$s^2 = SSE/n-k-1$
Total	n-1	$SST = y'y = Y'Y - n\bar{Y}^2$	$s_Y^2 = \frac{y'y}{n-1}$

The appropriate test is:

(3.2)

$$F = \frac{MSR}{s^2} = \frac{SSR/k}{SSE/n-k-1} = \frac{SSR}{SSE}\left(\frac{n-k-1}{k}\right) = \frac{SSR/y'y}{SSE/y'y}\left(\frac{n-k-1}{k}\right) = \frac{R^2}{1-R^2}\left(\frac{n-k-1}{k}\right)$$

There are k degrees of freedom associated with the mean square in the numerator of the F and n-k-1 degrees of freedom associated with the denominator. The degrees of the numerator and denominator components of the test.

The particular test presented in (3.2) provides additional information about the goodness of fit. The coefficient of determination indicates the proportion of variation in the endogenous variable accounted for by the exogenous variables in the regression equation

while the F test specified in (3.2) provides information about whether the amount of variation accounted for can be judged as being significantly different from zero at the selected level of significance. In this context being significantly different from zero means that not all elements of the β^{Δ} vector of regression parameters are zero. Hence, one or more of the exogenous variables are important in the sense that the variation in the one or more exogenous variables is significantly associated with variation in the endogenous variable.

Accounted for Variation and Deletion of Variables[2/]

At times it is of interest to know whether a given exogenous variable or subset of exogenous variables account for a significant amount of variation in the endogenous variable. Another way to view this is if a given exogenous variable or subset of exogenous variables were deleted from a regression model is the amount of variation accounted for by this subset of variables significantly different from zero? In other words one is seeking to determine if the parameters of one or more of the subset of exogenous variables are significantly different from zero.

Let us consider a regression equation which has g+1 exogenous variables; g explanatory variables plus a "dummy variable" which only takes on a value of one. The latter variable is included to obtain an intercept. g is considered to be less than k. The coefficients of this function could be estimated using least squares procedures and

---

2/ This also applies to addition of variables since in addition of variables one starts with a given model and then increases the number of exogenous variables included in the model while in deletion of variables the initial model has the largest number of exogenous variables which are reduced in subsequent specifications.

all quantities discussed thus far could be obtained for this specification. The main quantity of interest is the sum of squares regression which is:

(3.3) $SSR_g = b_g' X'Y - n\bar{Y}^2$ where b_g' are the least squares estimators for the relation with g+1 exogenous variables.

The difference between the sum squares when k+1 and g+1 exogenous variables are considered is:

(3.4) $$\Delta SSR = SSR - SSR_g$$

where SSR is given in (3.1). This is a difference between two sums of squares which is also a sum of squares. The sampling distribution of the difference between the two sums of squares is also a chi square.

The hypothesis to be tested is that the amount of variation in the endogenous variable accounted for by the (k-g) additional variables is not different from zero at some chosen level of significance. In terms of the parameters involved the null and alternative hypothesis are:

$$HO: \quad \beta_{g+1} = - - - = \beta_{k-1} = \beta_k = 0$$

HA: Not all $\beta_j = 0$ $j = g+1, \text{---}, k-1, k.$

Utilizing the information about the k exogenous variable model presented in (3.1), the appropriate test is

(3.5) $$F = \frac{\frac{\Delta SSR}{(k-g)}}{s^2}$$

with (k - g) degrees of freedom associated with the numerator and (n - k - 1) degrees of freedom associated with the denominator. The deletion of a single variable is a special case and (k - g) = 1 so that g = k - 1.

Illustration of Testing Addition to Sum of Squares Regression

Let us consider a hypothetical study of 89 firms in which we are interested in accounting for the variation in the endogenous variable, total cost (X_1) utilizing exogenous variables, rate of output (X_2) and rate of absenteeism (X_3).[3/] The means are

$$\bar{X}_1 = 5.8 \qquad \bar{X}_2 = 2.9 \qquad \bar{X}_3 = 3.9$$

and the matrix of moments about the means is

(3.6)

	X_1	X_2	X_3
X_1	113.6	36.8	39.1
X_2		50.5	-66.2
X_3			967.1

The problem is to estimate the relationship between X_1 and the other two variables, set up an analysis-of-variance table to show the reduction in the total sum of squares due to fitting X_2, the additional reduction due to fitting X_3, and the total reduction due to X_2 and X_3 together, and finally to test the overall effect and the partial X_3 effect, given X_2.

In estimating the coefficients of the relationship of interest in this case we need to concern ourselves with the lower right 2 by 2 submatrix of (3.6). The determinant needed in obtaining the inverse is:

(3.7)
$$\begin{vmatrix} 50.5 & -66.2 \\ -66.2 & 967.1 \end{vmatrix} = 44,456.11$$

3/ This is problem 5.9 in J. Johnston, Econometric Methods, Second Edition, 1972, p. 173.

The desired inverse is:

$$(3.8) \qquad [x'x]^{-1} = \frac{1}{44{,}456.11}\begin{bmatrix} 967.1 & 66.2 \\ 66.2 & 50.5 \end{bmatrix} = \begin{bmatrix} .02175 & .00149 \\ .00149 & .00114 \end{bmatrix}$$

and the least squares estimators expressed in matrix product form are:

$$(3.9) \qquad \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} .02175 & .00419 \\ .00149 & .00114 \end{bmatrix} \begin{bmatrix} 36.8 \\ 39.1 \end{bmatrix} = \begin{bmatrix} .85866 \\ .09941 \end{bmatrix}$$

hence:

$$b_1 = .85866$$

$$b_2 = .09941$$

Given the estimated values of the regression coefficients and the mean values of the variables the intercept may be estimated as follows:

$$(3.10) \qquad b_0 = \bar{X}_1 - b_1\bar{X}_2 - b_2\bar{X}_3$$

$$= 5.8 - (.85866)\ (2.9) - (.09941)\ (3.9)$$

$$b_0 = 2.92219$$

The estimated regression relationship (or estimating equation) is:

$$(3.11) \qquad X_1 = 2.922 + .859\ X_2 + .099\ X_3$$

The sum of squares regression when X_2 is the only exogenous variable i.e., simple linear regression is:

$$(3.12) \qquad SSR_1 = \frac{\Sigma x_1 x_2}{\Sigma x_2^2} \cdot \Sigma x_1 x_2 = \frac{(\Sigma x_1 x_2)^2}{\Sigma x_2^2} = \frac{(36.8)^2}{50.5} = 26.82$$

Note that $\frac{\Sigma x_1 x_2}{\Sigma x_2^2}$ is the least squares estimator of the regression coefficient in the simple linear regression framework.

The sum of squares regression due to X_2 and X_3 together that is when both X_2 and X_3 are incorporated as exogenous variables in the regression relationship is:

(3.13) $SSR_2 = b_1 \Sigma x_1 x_2 + b_2 \Sigma x_1 x_3 = .859\ (36.8) + .009\ (39.1) = 35.48$

Now from the information provided by (3.12) and (3.13) we can compute the additional amount of variation in X_1 which is accounted for by X_3 over and above the amount accounted for by X_2. Recognize that the increment of explained variation with which we are concerned is the additional amount which X_2 and X_3 together account for above what X_2 accounts for by itself when it is the only exogenous variable. Thus a specific sequence is being considered; X_2 first and then X_3. If the exogenous variables were orthogonal then the incremental sums of squares accounted for by each could be summed to obtain the sum of squares when all variables, in this case X_2 and X_3, are in the model. However, when exogenous variables are not orthogonal then the amount attributable to any one exogenous variable depends on which sequence is considered. In other words the amount of sum of squares regression attributable to X_3 alone or to X_3 and X_2 when X_3 is permitted to be incorporated into the equation ahead of (in terms of sequences) X_2. Now the additional sum of squares regression due to X_3 given that X_2 is already in the equation is simply the difference between (3.12) and (3.13).

(3.14) $\Delta SSR = SSR_2 - SSR_1 = 35.48 - 26.82 = 8.66$

The sum of squared residuals is:

(3.15) $SSE = \Sigma x_1^2 - SSR_2 = 113.6 - 35.48 = 78.12$

An estimate of the residual variance is:

(3.16) $s^2 = \frac{SSE}{n-3} = \frac{78.12}{86} = .9084$

The residual variance to be used in performing tests of hypotheses about the additional increment of sum of squares regression is the one presented in (3.16) in this case and in general the residual variance associated with the model containing the largest number of exogenous variables. The residual variance estimates associated with equations containing fewer exogenous variables contain the amounts of variation attributable to the exogenous variables not included in the models and hence these estimators are not as efficient estimators of the disturbance variance.

The analysis of variance table which gives the breakdown of the various sums of squares of interest and the corresponding mean squares is:

(3.17)

ANOV

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares
Due to X_2 and X_3	2	35.48	17.74
Due to X_2	1	26.82	26.82
Due to X_3 (above X_2)	1	8.66	8.66
Residual	86	78.18	.908
Total	88	113.60	

In testing the overall effect one is concerned about whether or not a significant amount of variation in the endogenous variable is accounted for by the set of exogenous variables used. Another way of saying this is that one is testing to determine if the β^{Δ}

vector is null, i.e., whether or not all the β_j excluding β_0 are equal to zero. Hence, the null and alternative hypotheses are:

(3.18) HO: $\beta\Delta = 0$

HA: $\beta^{\Delta} \neq 0$

or

HO: $\beta_j = 0$

HA: Not all $\beta_j = 0$ $\quad j = 2, 3, - - -, k$

The testing statistic is a ratio of two estimated variances which is an F.

$$(3.19) \qquad F = \frac{MSR}{s^2} = \frac{17.74}{.908} = 19.537^*$$

Choosing an α level of .05, $F_{(.05)(2,86)} = 3.11$ and hence the test indicates significance at the .05 level. At the .05 level (really at a higher level) the F-test indicates that the amount of total cost accounted for by the rate of output and rate of absenteeism, together is significant.

Testing the partial effect of X_3, given X_2 requires testing hypotheses about whether or not the additional amount of variation in X_1 accounted for by X_3 above that accounted by X_2 alone, is significantly different from zero. The specific null and alternative hypotheses are:

(3.20) HO: $\beta_3 = 0$

HA: $\beta_3 \neq 0$

The appropriate testing statistic is an F ratio:

$$(3.21) \qquad F = \frac{\Delta MSR}{s^2} = \frac{\frac{\Delta SSR}{1}}{s^2} = \frac{8.66}{.908} = 9.537^*$$

Choosing the .05 level of significance $F_{(.05)(1.86)} = 3.96$ and hence the computed F is in the rejection region. One concludes that the amount of variation in total cost accounted for by the rate of absenteeism given the rate of output, is significantly different from zero at the .05 level.

Note that the hypotheses presented in (3.20) and the test presented in (3.21) may be generalized at the addition or deletion of any g set of exogenous variables. This generalization is presented in (3.3) - (3.5).

Testing Hypotheses About Single Coefficients

In statistical relationships which are stochastic representations of some economic model, we are usually interested in testing hypotheses about individual coefficients to determine if they differ significantly, at a chosen level of significance, from zero or some specified nonzero value. The null and possible alternative hypotheses are:

HO:	$\beta_j = 0$		HO:	$\beta_j = K \neq 0$
HA_1	$\beta_j \neq 0$	or	HA_1	$\beta_j \neq K$
HA_2	$\beta_j < 0$		HA_2	$\beta_j < K$
HA_3	$\beta j > 0$		HA_3	$\beta_j > K$

The choice of the specific alternative hypothesis will depend upon what kind of prior information one has.

The appropriate test is

$$(3.22) \qquad t = \frac{b_j - \beta_j}{s_{b_j}}$$

where: b_j is the jth estimated coefficient

β_j is the jth parameter

s_{b_j} is the standard error of the jth estimated coefficient.

The degrees of freedom associated with t are the number associated with the standard error which is the same as the number of degrees of freedom associated with the residual variance. The degrees of freedom are $n - k - 1$. The standard error of b_j in the denominator is the square root of the variance:

$$s_{b_j} = \sqrt{c_{jj}\, s^2} \tag{3.23}$$

A t statistic which is significant at some specified level of significance indicates that the jth parameter is significantly different, significantly less, or significantly greater than zero or some specified nonzero value. Basically this indicates that the incremental changes in the endogenous variable associated with unit changes in the jth exogenous variable differ from zero at the level of significance being used.

The t test presented in (3.22) and the F presented in (3.5) in the special situation when $(k - g) = 1$ are two ways of testing whether the jth exogenous variable accounts for a significant amount of the variability in the endogenous variable. Except in the case where all exogenous variables are truly orthogonal these two tests are somewhat different. In the case of the t test all $k + 1$ exogenous variables were included in the model and the testing is done within the context of the (k+1) variable model. In the case of the F for differences in sum of squares regression, the jth variable is excluded from the (g+1) exogenous variable model but included in the (k+1) exogenous variable model. This latter test focuses on the contribution of the jth exogenous variable to the sum of squares regression. Because of nonzero correlation coefficients among the exogenous variables these tests may differ in terms of the conclusion one might make about the significance of the influences of the

Testing Hypotheses About Pairs of Coefficients

On occasions it may be of interest to perform tests of hypotheses between two coefficients in a given multiple regression equation. This may be accomplished by employing a t test. The hypotheses to be tested involve a relationship between two regression parameters. The null hypothesis may specify that the difference between two regression parameters is zero or some non-zero constant. The alternative hypothesis may specify that the difference between the two parameters is not equal to, less than, or greater than zero or a specified non-zero constant. The null and alternative hypotheses possibilities are:

HO:	$\beta_j - \beta_r = 0$		HO:	$\beta_j - \beta_r = K$
HA_1	$\beta_j - \beta_r \neq 0$		HA_1	$\beta_j - \beta_r \neq K$
HA_2	$\beta_j - \beta_r < 0$	or	HA_2	$\beta_j - \beta_r < K$
HA_3	$\beta_j - \beta_r > 0$		HA_3	$\beta_j - \beta_r > K$

Again the type of phenomena being investigated will dictate which null hypothesis is appropriate and the type of available prior information will determine which alternative hypothesis is appropriate.

The appropriate test is:

$$t = \frac{(b_j - b_r) - (\beta_j - \beta_r)}{s_{(b_j - b_r)}} \tag{3.24}$$

where: $(b_j - b_r)$ is the difference between the jth and rth estimated coefficients.

$(\beta_j - \beta_r)$ is the difference between the jth and rth regression parameters.

$s_{(b_j - b_r)}$ is the standard error of the difference between the jth and rth estimated regression coefficients.

The standard error of the difference may be derived from the covariance matrix of the coefficients and is:

$$(3.25) \qquad s_{(b_j - b_r)} = \sqrt{s^2 (c_{jj} + c_{rr} - 2c_{jr})}$$

For example the standard error of $(b_2 - b_3)$ is

$$s_{(b_2 - b_3)} = \sqrt{s^2 (c_{22} + c_{33} - 2c_{23})}$$

The degrees of freedom associated with the t test presented in (3.24) are $(n - k - 1)$. The degrees of freedom are those associated with the standard error which are the degrees of freedom for the residual variance.

A word of warning is needed in performing tests involving pairs of regression coefficients. The test is legitimate and appropriate for independent pairs of coefficients. Hence, it is not proper to test all possible pairs of coefficients of a set of coefficients. The same difficulty arises here as arises in testing pairs of means of a set of means. The test applies only to orthogonal pairs of coefficients.

Testing Linear Combinations of Regression Coefficients

The simpliest form of a test of linear combinations of regression coefficients was presented in (3.24). This entails the difference between two regression coefficients. Testing hypotheses about pairs of regression coefficients poses some interesting questions. One of these is that if we have $k > 2$ regression parameters, can bonefide tests be performed about all possible pairs? If not, then about which pairs or groups can tests be performed and how does one go about determining these?

The questions which have just been raised are clear and straightforward, unfortunately the answers or partial answers are not clear and straightforward. The t-test may be used to perform tests of significance about orthogonal or independent pairs of coefficients but not all pairs of all possible pairs of coefficients are orthogonal so that the t-test is not a legitimate test for all possible pairs. However, the t-test is appropriate for the orthogonal pairs. A sequential range test such as the Newman-Kuels' test,[4/] which was designed to test for differences among a set of sample means may be modified to be at least an approximate test for a set of regression coefficients. Focus here will be on some of the things which can and cannot be done in testing linear combinations of regression coefficients.

Let us consider a linear multiple regression equation of the form

(3.26) $\quad Y = X\beta + \varepsilon$

[4/] For additional details and tables of values, see Robert G. D. Steel and James H. Torrie, Principles and Procedures of Statistics, McGraw-Hill Book Company, 1960, pages 111-115 and Appendix Table A.8.

where:

Y is a (nx1) vector of values of the endogenous variable.

X is an n x (k+1) matrix of the values of the (k+1) predetermined variables (including the dummy variable for the intercept).

β is a (k+1) x 1 vector of the regression parameters.

ε is a n x 1 vector of distrubance terms.

The prediction or estimated equation corresponding to (3.26) where the estimated coefficients have been obtained by least squares techniques is in matrix form:

(3.27) $\hat{Y} = X\,b$

where

$\hat{Y}$ is a (nx1) vector of the predicted values of the endogenous variable.

b is the (k+1) x 1 vector of estimated regression coefficients including the intercept.

In non-matrix form the estimated equation is:

(3.28) $\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_kX_k$

An i subscript denoting the individual observations of the sample has been omitted and unless specified otherwise it is to be understood that sample size is n > (k+1).

Hypotheses about the individual coefficients may be tested using the t-test presented in (3.22). Specifically hypotheses about whether or not each of the β_j parameters are equal to zero or equal to some non-zero constant may be tested using the conventional t-test. Whether

a one or two tailed hypothesis is specified and a one or two tailed test is used depends on the nature of available prior information.

Now one can raise some interesting questions about whether β_1 is different from β_2, β_1 different from β_3, β_2 different from β_3 or any β_g different from β_h, $g \neq h$. These kinds of contrasts arise when the hypotheses being tested specify a relationship between two or more parameters. Significant differences between two parameters or among more than two parameters may be meaningful in an economic sense or in terms of some other discipline. Two approaches to answering questions about differences among parameters appear promising. One is to identify the orthogonal contrasts among the set of coefficients of interest and use the appropriate conventional tests to test hypotheses of interest. Second, is to modify a sequential range test for sets of means, recognizing that the b's are also a type of means, to apply to sets of estimated regression coefficients. The orthogonal contrasts will be considered first but before pursuing the topic some basic aspects of orthogonality and fundamentals of linear forms will be developed.

Orthogonal Comparisons

In terms of orthogonal comparisons, one is concerned with linear relations among sums such as

$$L = a_1^* S_1 + a_2^* S_2 + - - - + a_r^* S_r \tag{3.29}$$

where r is the number of variables, groups, or in the experimental sense treatments in the comparison and the a_j^* are coefficients of the sums making up the comparison.[5/] Let us assume that the number of

---

[5/] For additional information, see George Snedecor, Statistical Methods, Iowa State Press, 1956, p. 330 or Steele and Torrie, Principles and Procedures of Statistics, p. 216.

observations which go into making up each of the sums, S_j, is the same, say n the sample size. In experimental work, the number of observations on which each S is based is the number of replications but in the regression frameworks with which economists usually deal with this would be the number of observations or sample size.[6/]

Alternatively, the linear function in (3.4) may be expressed in terms of means:

$$(3.30) \quad L = a_1^* n \hat{u}_1 + a_2^* n \hat{u}_2 + - - - + a_r^* n \hat{u}_r$$

$$= n (a_1^* \hat{u}_1 + a_2^* \hat{u}_2 + - - - + a_r^* \hat{u}_r)$$

and

$$(3.31) \quad L = a_1 \hat{u}_1 + a_2 \hat{u}_2 + - - - + a_r \hat{u}_r$$

where $\hat{u}_j$ is an estimated mean corresponding to the sum S_j and $a_j = n a_j^*$.

There are two rules with regard to comparison which are relevant to assessing feasible contrasts among a set of regression coefficients. The first is that $L = \sum_{j=1}^{r} a_j \hat{u}_j$ or $L = \sum_{j=1}^{r} a_j^* S_j$ is a comparison if

$$(3.32) \quad \sum_{j=1}^{r} a_j^* = 0 \quad \text{and} \quad \sum_{j=1}^{r} a_j = 0$$

[6/] In analysis of variance of factorial arrangements of treatments where orthogonal comparisons are frequently of interest unequal numbers of observations among sums or means sometimes occur. This does not preclude making orthogonal comparisons; however, the unequal number of observations have to be taken into consideration. Instead of working with a_j^*'s or a_j's in the rules of orthogonal contrasts, $(n_j a_j^*)$'s are used and the orthogonal comparisons and corresponding tests are developed accordingly.

The second rule pertains to two comparisons say:

$$(3.33) \quad L_1 = a_{11}\hat{u}_1 + a_{12}\hat{u}_2 + a_{13}\hat{u}_3 + - - - + a_{1r}\hat{u}_r$$

$$(3.34) \quad L_2 = a_{21}\hat{u}_1 + a_{22}\hat{u}_2 + a_{23}\hat{u}_3 + - - - + a_{2r}\hat{u}_r$$

L_1 and L_2 are orthogonal if:

$$(3.35) \quad \sum_{j=1}^{r} a_{1j} = 0$$

$$(3.36) \quad \sum_{j=1}^{r} a_{2j} = 0$$

and

$$(3.37) \quad \sum_{j=1}^{r} a_{1j}a_{2j} = a_{11}a_{21} + a_{12}a_{22} + - - - + a_{1r}a_{2r} = 0$$

If a mean or sum does not enter into a particular comparison its coefficient is considered to be zero.

Now for illustrative purposes, let us consider a set of 5 parameters from (3.26) and attempt to identify which comparisons are orthogonal. We could choose any set and in practice this would be dictated by the questions we are attempting to answer which would indicate which ones are to be considered. For convenience purposes, let us just take the firat 5 - β_1, β_2, β_3, β_4 and β_5. In table form one set of orthogonal comparisons is:

Table 3.1. Orthogonal Comparisons Among Five Coefficients in Equation (3.26).

Comparison	β_1	β_2	β_3	β_4	β_5	Σa_{ij} and $\Sigma a_{ij}a_{kj}$
I. a_{1j}	+1	+1	+1	+1	-4	0
II. a_{2j}	-1	+1	0	0	0	0
a_{1j} a_{2j}	-1	+1	0	0	0	0
III. a_{3j}	0	0	+1	-1	0	0
a_{1j} a_{3j}	0	0	+1	-1	0	0
a_{2j} a_{3j}	0	0	0	0	0	0
IV. a_{4j}	+1	+1	-1	-1	0	0
a_{1j} a_{4j}	+1	+1	-1	-1	0	0
a_{2j} a_{4j}	-1	+1	0	0	0	0
a_{2j} a_{4j}	0	0	-1	+1	0	0

The comparisons presented in Table 1 are one set of four orthogonal comparisons among five selected coefficients of variables in (3.26). The five variables are comparable to 5 treatments and there are 4 degrees of freedom associated with the treatment affect and thus an orthogonal contrast for each individual degree of freedom. One thing which should be recognized is that there are alternative sets of four orthogonal comparisons which may be formulated for this same set of 5 coefficients. Which comparisons are considered in an orthogonal set of contrasts depends for which particular comparisons it is most important to have information. Which particular comparison or comparisons are formulated first in part determine which other orthogonal comparisons are possible given a set of coefficients or treatments. For example, an alternative set of orthogonal comparisons to those presented in Table 3.1 is:

(3.38)

	β_1	β_2	β_3	β_4	β_5
I	+1	+1	+1	+1	-4
II	+1	+1	+1	-3	0
III	+1	+1	-2	0	0
IV	+1	-1	0	0	0

The first comparison in Table 3.1 labeled I entails contrasting the first four coefficients against the 5th coefficient. For most sets of five parameters of continuous variables in regression equations this might not be a contrast of interest. However, if these were coefficients of discrete variables acting as surface shifters then this particular contrast might take quite different importance. The topic of zero-one variables in regression equation will be developed in another section

The second comparison in Table 3.1, II, is simply the difference between b_1 and b_2, and the null hypothesis to be tested is that $\beta_2 - \beta_1 = 0$. The t-test for pairs of regression coefficients presented in (3.22) is an appropriate test. Similarly the third orthogonal contrast entails the difference between b_3 and b_4 and the t-test presented in (3.22) is appropriate. Comaparison IV entails the sum of b_1 and b_2 contrasted with the sum of b_3 and b_4 or the average of b_1 and b_2 contrasted with the average of b_3 and b_4. The null hypothesis to be tested is that $(\beta_1 + \beta_2) - (\beta_3 + \beta_4) = 0$ or that $\frac{(\beta_1 + \beta_2)}{2} = \frac{(\beta_3 + \beta_4)}{2}$ or that $\frac{(\beta_1 + \beta_2)}{2} - \frac{(\beta_3 + \beta_4)}{2} = 0$. An appropriate test for this is not obvious and in fact we need to develop some fundamental concepts of linear forms before we can formulate the appropriate test. Intuitively one would feel that a modified version of the t might be appropriate. The form of the t which suggests itself from previous considerations is:

$$(3.39) \quad t = \frac{\left[\left(\frac{b_1 + b_2}{2} - \frac{b_3 + b_4}{2}\right)\right] - \left[\left(\frac{\beta_1 + \beta_2}{2} - \frac{\beta_3 + \beta_4}{2}\right)\right]}{s\left[\left(\frac{b_1 + b_2}{2} - \frac{b_3 + b_4}{2}\right)\right]}$$

The problem is how to get the standard error and to determine the appropriate degrees of freedom.

Some Fundamentals of Linear Forms

Let us consider the following linear form:

$$(3.40) \quad L = a_1b_1 + a_2b_2 + - - - - - + a_rb_r = \sum_{j=1}^{r} a_jb_j$$

where the a_j is some fixed constant and the b_j's are sample values which are not necessarily random. However, if the b_j's represent least squares estimators of regression coefficients they are random variables from a sampling viewpoint. They are a set of constants for any one sample of data but are random variables for different samples.

The expected value of the linear form given in (3.15) is:

$$(3.41) \qquad E(L) = \sum_{j=1}^{r} a_j E(b_j) = \sum_{j=1}^{r} a_j \beta_j$$

and the variance of the linear form is:

$$(3.42) \qquad \sigma_L^2 = E[L - (L)]^2 = E[\sum_{j=1}^{r} a_j(b_j - \beta_j)]^2$$

$$= \sum_{j=1}^{r} a_j^2 \sigma_{b_j}^2 + \sum_{j=1}^{r} \sum_{k=1}^{r} a_j a_k \sigma_{b_j b_k}$$

$$= \sum_{j=1}^{r} a_j^2 \sigma_{b_j}^2 + 2 \sum_{j<k} \sum a_j a_k \sigma_{b_j b_k}$$

where $\sigma_{b_j b_k}$ is the covariance between the j<u>th</u> and k<u>th</u> estimated regression coefficients.

If $\sigma_{b_j b_k} = 0$ then $\sigma_L^2 = \sum_{j=1}^{r} a_j^2 \sigma_{b_j}^2$

As an illustrative example let us consider the simple linear form:

$$(3.43) \qquad L = a_1 b_1 + a_2 b_2 = b_1 - b_2$$

where $a_1 = +1$ and $a_2 = -1$.

The variance of this linear form is:

$$(3.44) \qquad \sigma_L^2 = \sigma_{b_1 - b_2}^2 = (+1)^2 \sigma_{b_1}^2 + (-1)^2 \sigma_{b_2}^2 + (+1)(-1) \sigma_{b_1 b_2}$$

$$+ (-1)(+1) \sigma_{b_2 b_1}$$

$$= \sigma_{b_1}^2 + \sigma_{b_2}^2 - 2 \sigma_{b_1 b_2}$$

The estimated value of the variance given in (3.44) is:

$$(3.45) \qquad s^2_{b_1 - b_2} = s^2_{b_1} + s^2_{b_2} - 2\, s_{b_1 b_2}$$

$$= s^2(c_{11} + c_{22} - 2c_{12})$$

where s^2 is the estimated residual variance and c_{11}, c_{22}, c_{12} are elements of the $(X'X)^{-1}$ matrix. Note that (3.45) is the variance presented in (3.25) for the case where $j = 1$ and $r = 2$.

Now let us consider the linear form:

$$(3.46) \qquad L = a_1 b_1 + a_2 b_2 + a_3 b_3 + a_4 b_4$$

where $a_1 = a_2 = +1/2$ and $a_3 = a_4 = -1/2$. This is exactly the form under consideration in (3.14). Hence, the linear form of interest is:

$$(3.47) \qquad L = 1/2\, b_1 + 1/2\, b_2 - 1/2\, b_3 - 1/2\, b_4$$

The variance of the linear form in (3.47) is:

$$(3.48) \quad \sigma^2_L = (1/2)^2\sigma^2_{b_1} + (1/2)^2\sigma^2_{b_2} + (-1/2)^2\sigma^2_{b_3} + (-1/2)^2\sigma^2_{b_4} + 2(1/2)\,(1/2)$$

$$\sigma_{b_1 b_2} + (2)(1/2)(-1/2)\sigma_{b_1 b_3} + 2(1/2)(-1/2)\sigma_{b_1 b_4} + 2(1/2)(-1/2)$$

$$\sigma_{b_2 b_3} + 2(1/2)(-1/2)\sigma_{b_2 b_4} + 2(1/2)(-1/2)\sigma_{b_3 b_4}$$

$$= 1/4(\sigma^2_{b_1} + \sigma^2_{b_2} + \sigma^2_{b_3} + \sigma^2_{b_4}) + 1/2\, \sigma_{b_1 b_2} - 1/2\, \sigma_{b_1 b_3} - 1/2\, \sigma_{b_1 b_4}$$

$$- 1/2\, \sigma_{b_2 b_3} - 1/2\, \sigma_{b_2 b_4} + 1/2\, \sigma_{b_3 b_4}$$

The estimated value of (3.48) is:

$$(3.49)\quad s_L^2 = 1/4(s_{b_1}^2 + s_{b_2}^2 + s_{b_3}^2 + s_{b_4}^2) + 1/2\, s_{b_1b_2} - 1/2\, s_{b_1b_3} - 1/2\, s_{b_1b_4} - 1/2\, s_{b_2b_3} - 1/2\, s_{b_2b_4} + 1/2\, s_{b_3b_4}$$

$$= 1/4\, s^2(c_{11} + c_{22} + c_{33} + c_{44}) - 1/2\, s^2(-c_{12} + c_{13} + c_{14} + c_{23} + c_{24} - c_{34})$$

and hence the required standard error in (3.39) is:

$$(3.50)\quad s\left[\left(\frac{b_1 + b_2}{2} - \frac{b_3 + b_4}{2}\right)\right] = s[1/4(c_{11} + c_{22} + c_{33} + c_{44}) - 1/2\,(-c_{12} + c_{13} + c_{14} + c_{23} + c_{24} - c_{34})]^{1/2}$$

Thus the degrees of freedom associated with the standard error in (3.50) is the degrees of freedom associated with the estimated residual variance which is $(n - k - 1)$.

The hypotheses for testing comparison IV in Table I are:

$$HO: \left(\frac{\beta_1 + \beta_2}{2} - \frac{\beta_3 + \beta_4}{2}\right) = 0$$

$$HA: \left(\frac{\beta_1 + \beta_2}{2} \quad \frac{\beta_3 + \beta_4}{2}\right) \neq 0$$

The appropriate test is:

$$(3.51)\quad t = \frac{\left(\frac{b_1 + b_2}{2}\right) - \left(\frac{b_3 + b_4}{2}\right)}{s[1/4(c_{11} + c_{22} + c_{33} + c_{44}) - 1/2\,(-c_{12}+c_{13}+c_{14}+c_{23}+c_{24}-c_{34})]^{1/2}}$$

with $(n - k - 1)$ degrees of freedom, the same degree of freedom associated with the residual variance.

Note that the particular values of a_j's used in (3.47) are arbitrary but they must sum to zero. If we do some factoring in (3.51) we can express the t as:

$$(3.52)\quad t = \frac{(b_1 + b_2) - (b_3 + b_4)}{s[(c_{11} + c_{22} + c_{33} + c_{44}) - 2(-c_{12}+c_{13}+c_{14}+c_{23}+c_{24}-c_{34})]^{1/2}}$$

Also, a F test could have been used to test the hypotheses since $t^2 = F$ with one degree of freedom associated with the numerator and (n-k-1) degrees of freedom associated with the denominator. This test would simply be:

$$(3.53)\quad F = \frac{[(b_1 + b_2) - (b_3 + b_4)]^2}{s^2[(c_{11}+c_{22}+c_{33}+c_{44}) - 2(-c_{12}+c_{13}+c_{14}+c_{23}+c_{24}-c_{34})]}$$

Now we can briefly consider the first comparison (I in Table 3.1) which we indicated would probably not be of interest in most regression problems. It entailed comparing the sum of four regression parameters with the fifth regression parameter. If for some reason we are interested in the sum or mean of four parameters versus the fifth one, we may develop the appropriate test along the lines followed to develop the test for $(\beta_1 + \beta_2)$ versus $(\beta_3 + \beta_4)$. The hypotheses to be tested are:

$$HO: \beta_1 + \beta_2 + \beta_3 + \beta_4 = \beta_5$$

or

$$HO: \beta_1 + \beta_2 + \beta_3 + \beta_4 - \beta_5 = 0$$

$$HA: \beta_1 + \beta_2 + \beta_3 + \beta_4 \neq \beta_5$$

The linear form of interest is:

$$(3.54)\quad L_1 = a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4 + a_5b_5$$

where $a_1 = a_2 = a_3 = a_4 = 1$ and $a_5 = -4$.

Hence,

$$(3.55) \qquad L_1 = b_1 + b_2 + b_4 - 4b_5$$

The variance of the linear form in (3.55) is:

$$(3.56) \quad \sigma^2_{L_1} = \sigma^2_{b_1} + \sigma^2_{b_2} + \sigma^2_{b_3} + \sigma^2_{b_4} + (-4)^2 \sigma^2_{b_5} + 2\sigma_{b_1b_2} + 2\sigma_{b_1b_3} +$$

$$2\sigma_{b_1b_4} + 2(1)(-4)\sigma_{b_1b_5} + 2\sigma_{b_2b_3} + 2\sigma_{b_2b_4} + 2(1)(-4)\sigma_{b_2b_5}$$

$$+ 2\sigma_{b_3b_4} + 2(1)(-4)\sigma_{b_3b_5} + 2(1)(-4)\sigma_{b_4b_5}$$

$$= \sigma^2_{b_1} + \sigma^2_{b_2} + \sigma^2_{b_3} + \sigma^2_{b_4} + 16\sigma^2_{b_5} + 2[\sigma_{b_1b_2} + \sigma_{b_1b_3} + \sigma_{b_1b_4}$$

$$- 4\sigma_{b_1b_5} + \sigma_{b_2b_3} + \sigma_{b_2b_4} - 4\sigma_{b_2b_5} + \sigma_{b_3b_4} - 4\sigma_{b_3b_5} - 4\sigma_{b_4b_5}]$$

The estimated value of (3.56) is:

$$(3.57) \quad s^2_{L_1} = s^2_{b_1} + s^2_{b_2} + s^2_{b_3} + s^2_{b_4} + 16s^2_{b_5} + 2[s_{b_1b_2} + s_{b_1b_3} + s_{b_1b_4}$$

$$- 4s_{b_1b_5} + s_{b_2b_3} + s_{b_2b_4} - 4s_{b_2b_5} + s_{b_3b_4} - 4s_{b_3b_5} - 4s_{b_4b_5}]$$

$$= s^2(c_{11} + c_{22} + c_{33} + c_{44} + 16c_{55}) + 2s^2[c_{12} + c_{13} + c_{14}$$

$$- 4c_{15} + c_{23} + c_{24} - 4c_{25} + c_{34} - 4c_{35} - 4c_{45}]$$

The respective standard error of L_1 is:

$$(3.58) \quad s_{L_1} = \sqrt{s^2_{L_1}} = [s^2(c_{11}+c_{22}+c_{33}+c_{44}+16c_{55}) + 2s^2(c_{12}+c_{13}+c_{14}-4c_{15}$$

$$+ c_{23}+c_{24}-4c_{25}+c_{34}-4c_{35}-c_{45})]^{1/2}$$

In (3.57) and (3.58) the c_{ij}'s are elements of $(X'X)^{-1}$ used in obtaining b in (3.27).

Again we are really interested in testing the difference between two means where one of these means is really a mean of 4 means, i.e., 1/4 ($\beta_1 + \beta_2 + \beta_3 + \beta_4$). A suitable test is the t-test as follows:

$$(3.59) \quad t = \frac{4\left[\left(\frac{b_1 + b_2 + b_3 + b_4}{4}\right) - b_5\right]}{s_{L_1}}$$

$$= \frac{4\left[\left(\frac{b_1 + b_2 + b_3 + b_4}{4}\right) - b_5\right]}{4s\left\{\frac{1}{16}\left[(c_{11}+c_{22}+c_{33}+c_{44}+16c_{55}) + 2(c_{12}+c_{13}+c_{14}-4c_{15}+c_{23} -4c_{25}+c_{34} - 4c_{35}-4c_{45})\right]\right\}^{1/2}}$$

$$= \frac{\frac{b_1 + b_2 + b_3 + b_4}{4} - b_5}{s\left\{\frac{1}{16}\left[(c_{11}+c_{22}+c_{33}+c_{44}+16c_{55}) + 2(c_{12}+c_{13}+c_{14}-4c_{15}+c_{23} +c_{24}-4c_{25}+c_{34}-4c_{35}-4c_{45})\right]\right\}^{1/2}}$$

with (n-k-1) degrees of freedom. As before the F-test is an alternative test which could be used to test the hypotheses. We are dealing with one degree of freedom in every one of these orthogonal contrasts and $t^2 = F$ with one degree of freedom associated with the numerator and (n-k-1) degrees of freedom associated with the denominator. The f is of the form:

$$(3.60) \quad F = t^2 = \frac{\left[\frac{b_1 + b_2 + b_3 + b_4}{4} - b_5\right]^2}{s^2\left(\frac{1}{16}\right)\left[(c_{11}+c_{22}+c_{33}+c_{44}+16c_{55})+2(c_{12}+c_{13}+c_{14}-4c_{15}+c_{23} +c_{24}-4c_{25}+c_{34}-4c_{35}-4c_{45}\right]}$$

Summary of Orthogonal Comparisons and Linear Forms

By using some of the basic concepts of orthogonal contrasts and linear forms the possible tests among five coefficients of variables in a regression equation evolved as various versions of the t-tests. The second and third comparisons involved using the t-test for difference between pairs of coefficients, and the first and fourth involved a linear form with several components but the basic test was still a simple t-test. In the first and fourth comparisons one of the more difficult tasks was that of obtaining the appropriate standard error. The F-test is an alternative test in all of these comparisons since $t^2 = F$ with one degree of freedom associated with the numerator and (n-k-1) degrees of freedom associated with the denominator.

Even though only 5 parameters were used for illustrative purposes it should be clear that the methods apply to any number of coefficients. Of course, the various orthogonal comparisons become more complex as the number of coefficients under consideration increases. However, not all comparisons of an orthogonal set of contrasts need to be submitted to statistical tests. Often a few selected comparisons provide the desired information if the particular comparisons are selected properly. In problems with large numbers of coefficients, the selection of orthogonal comparisons to provide the desired information becomes one of the key parts of the analysis.

Newman-Kuels Test

The difficulties with performing tests about all possible pairs of means and hence making non-orthogonal comparisons has long been recognized in the testing of sets of sample means. The basic problem with

testing all possible pairs is that the level of significance decreases as non-orthogonal comparisons. In other words one may be performing the tests at some chosen level of significance say α and using rejection values for that level of significance when in fact the true level of significance may be considerably less. As the number of coefficients or means in the set increases the loss in the level of significance becomes more severe. The outcome is that too many differences are judged to be significantly different from zero at a chosen probability level when in fact the differences are not significantly different from zero.

The Newman-Kuels test is a sequential range test which is designed to permit testing a wide range of possible pairs. The basic notion underlying this test is that the ranges of differences which are specified as significant at a chosen level of significance are systematically adjusted depending upon the number of coefficients or means in the particular group being tested so as to offset the loss of the level of significance. In fact the ranges or differences which would be judged significant are varied accordingly.

In performing a Newman-Kuels test the first step is to arrange the estimated coefficients in decreasing order of magnitude such that:

(3.61) $b_1 > b_2 > b_3 - - - - > b_h$

where the subscript i refers to the rank position of the coefficient. Next compute the differences between coefficients and prepare these in a table. The differences are highest and lowest, highest and second lowest, second highest and lowest, second highest and second lowest, etc., etc. It is easier to visualize and work with these differences if they are presented in a table such as the following:

Table 3.2. Differences Among Estimated Regression Coefficients from Highest to Lowest.

Rank	h	h-1	- - - - - - -	3	2
1	$b_1 - b_h$	$b_1 - b_{h-1}$	- - - - - - -	$b_1 - b_3$	$b_1 - b_2$
2	$b_2 - b_h$	$b_2 - b_{h-1}$	- - - - - - -	$b_2 - b_3$	
.					
.					
.					
.					
h-2	$b_{h-2} - b_h$	$b_{h-2} - b_{h-1}$			
h-1	$b_{h-1} - b_h$				

The next step is to prepare a list of least significant ranges. The ranges are as follow:

$$R_{h(j,k)} = q_\alpha(h,\ n-k-1) \cdot \frac{s_{(b_j - b_k)}}{\sqrt{2}} \qquad (3.62)$$ 7/

where $R_{h(j,k)}$ is the least significant range for the comparison involving h coefficients and specifically for the difference between b_j and b_k.

$q_\alpha(h,\ n-k-1)$ is the value of the special studentized range at the α level of significance and for h coefficients and n-k-1 degrees of freedom.

7/ The $\frac{1}{\sqrt{2}}$ is needed because the $q_\alpha(h,\ n-k-1)$ factor from the special Table A.8 on page 444 of Steele and Torrie, 1960, encompasses a $\sqrt{2}$ factor due to the fact that it is developed for the standard error of the mean difference with common variance and equal sample size and hence

$$s_{\bar{X}_j - \bar{X}_k} = \sqrt{\frac{s^2}{n} + \frac{s^2}{n}} = \sqrt{\frac{2s^2}{n}} = \frac{s}{\sqrt{n}}\sqrt{2} = s_{\bar{X}}\sqrt{2}$$

α is the level of significance and generally the studentized values are available for $\alpha = .05$ and $\alpha = .01$.

$s_{(b_j - b_k)}$ is the standard error between the jth and kth estimated regression coefficients. $s_{b_j - b_r} = \sqrt{s^2(c_{jj} + c_{kk} - 2c_{jk})}$

The least significant range values are:

$$R_{h(1,h)} = q_\alpha(h,\ n-k-1)\ \frac{s_{(b_1 - b_h)}}{\sqrt{2}}$$

$$R_{h-1\ (1,h-1)} = q_\alpha(h-1,\ n-k-1)\ \frac{s_{(b_1 - b_{h-1})}}{\sqrt{2}}$$

$$\vdots$$

$$R_{2(1,2)} = q_\alpha(2,n-k-1)\ \frac{s_{(b_1 - b_2)}}{\sqrt{2}}$$

$$R_{h-1}(2,h) = q_\alpha\ (h-1,\ n-k-1)\ \frac{s_{(b_2 - b_h)}}{\sqrt{2}}$$

$$R_{h-2}(2,h-1) = q_\alpha\ (h-2,\ n-k-1)\ \frac{s_{(b_2 - b_{h-1})}}{\sqrt{2}}$$

$$\vdots$$

$$R_{2(2,3)} = q_\alpha\ (2,\ n-k-1)\ \frac{s_{(b_2 - b_3)}}{\sqrt{2}}$$

$$\vdots$$

$$R_2(h-1,\ h) = q_\alpha\ (2,\ n-k-1)\ \frac{s_{(b_{h-1} - \delta_k)}}{\sqrt{2}}$$

The test proceeds sequentially by rows, starting with row 1 of Table 3.2.

1. Declare $b_1 - b_h$ significant if it is greater than $R_{h(1,k)}$. If not, stop. If it is significant, proceed.

2. Declare $b_1 - b_{h-1}$ significant if it is greater than $R_{h-1(1,h-1)}$. If not, stop. If it is significant, proceed.

3. Declare $b_1 - b_{h-2}$ significant if it is greater than $R_{h-2(1,h-2)}$. If not, stop. If it is significant, proceed.

Continue in this way until the first non-significant difference occurs in Row 1. All differences in coefficients to the <u>right of</u> this non-significant difference are also declared non-significant. Also, all differences <u>below and to the right of the non-significant differences in Row 1 are also declared non-significant</u>.

Now proceed to row 2 of Table 3.2 but <u>consider only those columns whose row differences were declared significant</u>.

4. Declare $b_2 - b_h$ significant if it is greater than $R_{h-2(2,h)}$. If not, stop. If it is significant, proceed.

5. Declare $b_2 - b_{n-1}$ significant if it is greater than $R_{h-3(2,h-1)}$. If not, stop. If it is significant, proceed.

Continue in this way until the first non-significant difference occurs. All differences <u>to the right</u> of this in row 2 are also declared non-significant. All differences <u>below</u> these non-significant differences in row 2, in the same columns, are also declared non-significant.

Now we proceed to row 3 of Table 2 considering only those columns whose row 2 and hence also row 1 entries were significant.

6. Declare $b_3 - b_h$ significant if it is greater than $R_{h-3(3,h)}$. If not, stop. If it is significant, proceed.

Continue in this manner until a non-significant difference is obtained in column h of Table 3.2 in which case the testing would be terminated or until the final test in row(h-1) is performed. This test permits assessing all possible pairs of a set of coefficients and it is fairly easy to perform. Usually there are some differences for which the test results are not clear cut and one is not able to say as much about them as one would like.

Finally, the Newman-Kuels test is designed to overcome the problem of the changing level of significance when conventional statistical tests for testing differences between pairs of parameters are applied to sets of non-orthogonal differences between pairs of parameters. For the Newman-Kuels test the α level is the level of significance and denotes the probability of committing a Type I error (probability of rejecting a null hypothesis which in fact is true) for any one testing event given the particular test being used. For the Newman-Kuels test the α refers to the probability of committing at least one Type I error among the $\frac{h(h-1)}{2}$ differences tested in Table 3.2 for infinite degrees of freedom if in fact all null hypotheses are true, i.e. $\beta_1=\beta_2= \text{---} = \beta_h$. In fact the α really represents the probability of committing a Type I error on the largest difference for infinite degrees if in fact all null hypotheses are true.

Equality Between Coefficients of Two or More Relations

A problem which often arises in estimation and hypothesis testing is whether or not sets of data should be pooled and one regression estimated or whether a separate regression equation should be estimated for each group of data.[8/] This type of problem arises when basically the same type of relationship is being estimated for several geographical regions, several time periods, or several classes of any number of qualitative variables.

Let us suppose that

k = number of predetermined variables[9/] $j = 1, 2, \ldots, k$

g = number of groups each with some n_i observations.

The relationship we are concerned with are

$$
\begin{aligned}
Y_1 &= X_1\beta_1 + \varepsilon_1 \\
Y_1 &= X_2\ _2 + \varepsilon_2 \\
&\vdots \\
Y_g &= X_g\beta_g + \varepsilon_g
\end{aligned}
\tag{3.63}
$$

[8/] Additional information about analyzing various aspects of equality among regression relations is presented in J. Johnston, Econometric Methods, Second Edition, 1972, pp. 192-207 and M. H. Quenouille, Associated Measurements, Academic Press, Inc. Publishers, 1952, pp. 75-79.

[9/] Predetermined variables include exogenous variables and lagged endogenous variables.

and

$$Y_T = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ Y_g \end{bmatrix} \tag{3.64}$$

and

$$X_T = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ X_g \end{bmatrix} \tag{3.65}$$

and

$$\varepsilon_T \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_g \end{bmatrix} \tag{3.66}$$

Y_T, X_T, and ε_T simply refer to the pooled observations from all g groups. The relationship pertaining to all the data pooled together is:

$$Y_T = X_T\beta_T + \varepsilon_T \tag{3.67}$$

The number of observations in Y_T, X_T and ε_T is $\sum_{i=1}^{g} n_i$. The number of observations in the g groups need not be the same; however, we are assuming that the number of predetermined variables in the relationships for each of the groups is the same: in fact, we are assuming that we are considering the same predetermined variables in each group but of course, the values

these predetermined variables take on may vary from group to group.

Now the question that one is concerned with in testing the equality of regression coefficients is whether $\beta_1, \beta_2 \cdot \ldots \beta_g$ are different from each other or whether they may be viewed as a homogeneous set of parameters. If the sets of regression parameters are not different from each other then there is a common set of parameters for all groups and that common set is β_T. If β_T is the appropriate vector of parameters then we want the estimators to be based on all of the available data pooled together. On the other hand, if there are differences among the sets of parameters, $\beta_1, \beta_2, - - - \beta_g$ then we will want the estimators of each of these sets to be based only on the data to which the set pertains. To perform the tests needed to determine which of the sets of parameters are appropriate estimators will have to be obtained for each of the groups separately and for the total of all groups when the data are pooled together.

Consider the normal equations for each group

$$
\begin{aligned}
&(X_1' X_1)\ b_1 = X_1' Y_1 \\
&(X_2' X_2)\ b_2 = X_2' Y_2 \\
&\quad\vdots \qquad\qquad \vdots \\
&(X_g' X_g)\ b_g = X_g' Y_g
\end{aligned}
\tag{3.68}
$$

where $b_1 - - - b_g$ are the least squares estimators for $\beta_1 - - - \beta_g$. $X_1 - - - X_g$ each contain a vector of ones so that intercepts may be obtained for each. The variation in the endogenous variable for any of the g group is:

$$
y_i' y_i = Y_i' Y_i - n_i \bar{Y}_i^2 \qquad i = 1,2, \ldots, g \tag{3.69}
$$

The sum of the variation over all of the g groups when each is viewed separately is:

$$(3.70)\quad \sum_{i=1}^{g} y_i' y_i = \sum_{i=1}^{g} Y_i' Y_i - \sum_{i=1}^{g} n_i \bar{Y}_i^2 = y_1' y_1 + y_2' y_2 + - - - - y_g' y_g$$

The amount of variation accounted for by the regression for any one of the g groups is:

$$(3.71)\quad SSR_i = b_i X_i Y_i - n_i \bar{Y}_i^2 \qquad i = 1, 2, - - -, g$$

The sum of squares regression over all of the g groups when each is viewed separately is:

$$(3.72)\quad \sum_{i=1}^{g} SSR_i = \sum_{i=1}^{g} b_i' X_i Y_i - \sum_{i=1}^{g} n_i Y_i^2$$

The sum of squared residual for any one of the g groups is:

$$(3.73)\quad SSE_i = e_i' e_i = (Y_i - X_i b_i)' (Y_i - X_i b_i)$$

The sum of squared residuals over all of the g groups when each is viewed separately is:

$$(3.74)\quad \sum_{i=1}^{g} SSE_i = \sum_{i=1}^{g} e_i' e_i$$

The information about the individual regressions fit to each group may be conveniently summarized in an analysis of variance table.

SOURCE	DEGREES OF FREEDOM	SUMS OF SQUARES
Due to regression	$g \quad k$	$\sum_{i=1}^{g} b_1' X_i' Y_i - \sum_{i=1}^{n} n_i \bar{Y}^2 = \sum_{i=1}^{g} SSR_i = S_2$
Dev. from regression	$\sum_{i=1}^{g} n_i - gk - g$	$\sum_{i=1}^{g} e_i' e_i = \sum_{i=1}^{g} SSE_i = S_1$
Total	$\sum_{i=1}^{g} n_i - g$	$\sum_{i=1}^{g} y_i' y_i$

Now let us focus on the regression relationship when all the data from the g groups are pooled together. The pooled data are presented in (3.64)-(3.66) and the relationship of interest is given in (3.67). The normal equations are:

$$(3.76) \quad (X_T' X_T)\, b_T = X_T' Y_T$$

and

$$(3.77) \quad b_T = (X_T' X_T)^{-1} X_T' Y_T$$

The variation in the endogenous variable is:

$$(3.78) \quad y_T' y_T = Y_T' Y_T - \sum_{i=1}^{g} n_i \bar{Y}_i^2$$

The sum of squares regression is:

$$(3.79) \quad SSR_T = b_T' X_T' Y_T - \sum_{i=1}^{g} n_i \bar{Y}_T^2$$

The sum of squared residuals is:

$$(3.80) \quad SSE_T = e_T' e_T = (Y_T - X_T b_T)' (Y_T - X_T b_T)$$

The information about the regression relationship estimated from the pooled data may be conveniently summarized in an analysis of variance table as follows

Source	Degrees of Freedom	Sum of Squares
Due to regression	k	$b_T' X_T' Y_T - \sum_{i=1}^{g} n_i \bar{Y}_T^2 = SSR_T$
Dev. from regression	$\sum_{i=1}^{g} n_i - k - 1$	$e_T' e_T = SSE_T = S_T$
Total	$\sum_{i=1}^{g} n_i - 1$	$y_T' y_T$

The test for aggregation is a test of the following hypotheses·

(3.82) HO: One regression relationship for all groups.

HA· Not one regression relationship for all groups.

The appropriate test is the F ratio.

$$F = \frac{(S_T - S_1) \,/\, (k+1)(g-1)}{S_1 \,/\, \sum_{i=1}^{g} n_i - gk - g} \tag{3.83}$$

with $(k+1)(g-1)$ degrees of freedom in the numerator and $\sum_{i=1}^{g} n_i - gk - g$ degrees of freedom in the denominator. $(S_T - S_1)$ is a measure of between group variability and we have the between group variability for each coefficient and the mean. For each coefficient and the mean, we have (g-1) degrees of freedom. There are k coefficients plus the mean thus the degrees of freedom for the numerator are (k+1) (g-1) which is also the difference between the degrees of freedom for S_T and the degrees of freedom for S_1.

If the test in (3.83) turns out to be nonsignificant at the chosen probability level, the data may be aggregated and the relationship estimated from the pooled data is appropriate. If the test in (3.83) indicates significance at the chosen probability level, the data should not be aggregated.

If the test for aggregation indicates that the data should not be aggregated one may wish to further analyze the situation to determine if the slope coefficients or intercepts are responsible for the heterogeneity. In order to test hypotheses about the slope coefficients we need to estimate a regression relationship based on the sum of the normal equations expressed in moments about the means. This basically provides us with information about whether or not a regression fit to group means (Y means on X means) is linear or not. We will want to adjust the sum of squares used in testing hypotheses about the slope coefficients for the effects of a relationship among the group means.

Let us now sum the normal equations expressed in moments about the means.

$$(3.84) \quad (x_s' x_s) = (x_1' x_1) + (x_2' x_2) + - - - - + (x_g' x_g)$$

and

$$(3.85) \quad (x_s' y_s) = (x_1' y_1) + (x_2' y_2) + - - - - + x_g' y_g$$

where x_i and y_i represent deviations from their respective means.

$$(3.86) \quad (x_s' x_s)\, b_s = x_s' y_s$$

and

$$(3.87) \quad b_s = (x_s' x_s)^{-1} x_s' y_s$$

where b_s are the least squares estimators based on the information in the sums of the normal equations. The sum of squares regression based on the summed normal equation is:

(3.88) $= SSR_s = b_s' x_s' y_s$

Now to investigate whether or not the slope parameters differ we test the hypotheses.

(3.89) HO: $\beta_1 = \beta_2 = - - - = \beta_g$ (slope parameters equal)

HA: Not all slope parameters are equal.

The appropriate test is the following F:

$$(3.90) \qquad F = \frac{(S_2 - S_3) \;/\; k(g-1)}{S_1 \;/\; \sum_{i=1}^{g} n_i - gk - g}$$

with k (g-1) degrees of freedom in the numerator and $\sum_{i=1}^{g} n_i - gk - g$ degrees of freedom in the denominator. If (3.90) indicates significance at the chosen probability level, we conclude that the reason we could not aggregate is because the slope coefficients were different. If this test is not significant, we proceed to a test about differences in intercepts. In fact if /the first test is significant and test of hypotheses about β's is not then the test of hypotheses about intercepts has to be significant.

In investigating differences in intercepts we test the following hypotheses

(3.91) HO: Intercepts are equal

HA: Not all intercepts are equal

The appropriate test is:

$$(3.92) \qquad F = \frac{(S_T - S_2 + S_3 - S_1) \;/\; g-1}{S_1 \;/\; \sum_{i=1}^{g} n_i - gk - g}$$

with (g-1) degrees of freedom in the numerator and $\Sigma\, n_i - gk - g$ degrees of freedom in the denominator. If (3.92) indicates significance at the chosen probability level, then we conclude that the reason we should not have aggregated is that the intercepts are different among the groups.

Again, the information about the sum of squares pertaining to the tests about the slope coefficients and the intercept may be conveniently summarized in a composite analysis of variance table.

(3.93)

Source	Degree of Freedom	Sum of Squares
Due to one reg^{n} to all data	k	$SSR_2 = b_T' X_T' Y_T - \sum_{i=1}^{g} n_i \bar{Y}_T^2$
Bet. group reg^{n} slope coefficients	$k(g-1)$	$S_2 - S_3$
Bet. group reg^{n} Intercepts	$g - 1$	$S_T - (S_2 - S_3) - S_1$
Residual	$\sum_{i=1}^{g} n_i - gk - g$	S_1
Total	$\sum_{i=1}^{g} n_i - 1$	$y_T' y_T$

There are other procedures for testing hypotheses about equality of coefficients among two or more linear regression relationships. Discrete variables which take on only values of one or zero sometimes referred to as "dummy" variables is a method of generalizing linear relationships which has gained considerable popularity in recent years. The zero-one variables may be used to allow for intercept differences, slope differences or both intercept and slope differences among linear regression relationships. The next chapter is devoted to the various uses of zero-variables and appropriate estimation and testing procedures.

Chapter 4: Use of Zero-One Variables in Linear Regressions

Introduction

The use of zero-one variables, sometimes referred to as "dummy" variables, has become prevalent in regression models representing economic phenomena. In these types of models the zero-one variables for which continuous measurements are either not available or not possible. The zero-one variables may be used to represent temporal effects such as seasons (months, quarters, etc.) wartime and peace time years, different political regimes, or different governmental programs. In some relationships the zero-one variables may represent different geographical regions, different neighborhoods, different types of farming regions or different metropolitan areas. Many types of qualitative variables such as sex, marital status, occupational classes, ethnic background, race, and quantitative variables such as age and income groupings may be included in a model using zero-one variables.

Most frequently when zero-one variables occur in a regression model they are predetermined variables. However, the phenomena represented by the endogenous variable may occasionally be of the zero-one form. When the endogenous variable is of zero-one type the predetermined may be either continuous, zero-one or a mixture of zero-one and continuous variables.

Zero-one variables are basically "yes-no" types of variables. They represent the occurrence or non-occurrence of a particular attribute. An attribute being represented by zero-one variables may consist of two more categories but these categories must be mutually exclusive. In other words if an attribute belongs to one category it cannot concurrently belong to any of the other categories. For example, if we consider four

seasons then a particular observation pertains to one of these seasons but only one. When using zero-one variables, the categories have to be established so that they are mutually exclusive or otherwise zero-one variables are not appropriate.

As their name implies, zero-one variables take on a value of either one or zero, depending on the occurence or non-occurence of a particular attribute. For models in logarithmic form, the discrete variables have to take on values such that the logarithms of the discrete variables are zero or one. Generally, if the attribute occurs, the variable takes on a value of one. However, there is no hard, fast rule that dictates this. In terms of the mechanics of using zero-one variables, the variable could take on a value of zero when the attribute occurs and a value of one when the attribute does not occur. In some cases one formulation may be more convenient for interpretation and subsequent analyses while in other cases the other formulation may be more suitable. Ease and convenience of analysis should guide the construction of the zero-one variable.

Zero-one variables incorporated in a linear multiple regression equation allow for differences in intercepts or differences of slopes of various dimensions of the surface or a combination of differences in intercepts and slopes. To allow for differences in slopes of various dimensions of the surface pertaining to various continuous variables involves generating new variables which are products of the zero-one variables and the continuous variables of interest. These new variables represent the interactions between the zero-one variables and the particular continuous variables. Significant differences among intercepts and/or among slopes generally have economic meaning (or other discipline meaning) and hence the tendency is to want to test hypotheses about all differences among intercepts and among slopes in order to conclude just which ones are different from which others. However, as with any other set of regression parameters problems arise in performing tests for various contrasts and all possible pairs of differences. If non - orthogonal

contrasts are going to be submitted to statistical tests then tests designed for such comparisons should be used.

Before illustrating the use of zero-one variables the key purpose of the use of zero-one variables merits emphasis The key purpose is to achieve a greater degree of generalization of a model. Rather than attempting to estimate a model for each attribute or sets of attributes, which might not only be impractical but impossible, one general model is estimated and the zero-one variables represent the effects of the different attributes which are assumed to account for the differences in models among the various attributes. From an estimation viewpoint all data are grouped together and the zero-one variables are supposed to represent the differences among categories or class of the qualitative or categorized quantitative variables.

Zero-One Variables as Intercept Shifters

Let us suppose we are concerned with a linear multiple regression model where one of the attributes to be taken into consideration consists of (h+1) categories or groups to be represented by zero-one variables. The categories must be mutually exclusive and must include all possible values that the attribute can assume. Let Y be a continuous quantitative variable and in addition to the zero-one variables there are other predetermined variables which are continuous quantitative variables.

Specifically let us consider a case where (h+1) = 5. There are five categories or classes of the attribute which may be represented by five zero-one variables: D_1, D_2, D_3, D_4 and D_5. The discrete variables take on values in the following way. D_1 takes on a value of one if the observation pertains to the first category; otherwise it takes on a value of zero. D_2 takes on a value of one when the observation pertains to the second category; otherwise it takes on a value of zero. D_3, D_4 and D_5 take on values on a similar manner. Hence, every observation in the

sample must pertain to one of the five categories and each observation pertains to only one of the categories. The number of observations pertaining to each category does not have to be equal. However, there must be at least two observations in each category or degrees of freedom problems will arise in estimation. Furthermore, when the numbers in any category are very small large sampling variation associated with very small samples may occur. The number of ones in each category are the number of replications of the effect which the category represents.

The observations for the five zero-one variables which would comprise a partition of the X matrix are given in Table 4.1. The observations in Table 4.1 are arrayed in a systematic fashion such that observation 1 pertained to category 1, observation 2 to category 2, etc., and observation 6 to category 1 again, etc. Obviously in reality the observations need not occur in this manner and probably would not, but by shuffling observations they could be arranged in this manner.

There is one very important thing to note about the observations in Table 4.1. The sum across columns for any row is equal to 1. Hence,

$$(4.1) \qquad \sum_{j=1}^{5} D_j = [1]_{nx1} = X_o$$

The sum of all zero-one variables of a set equals a column vector of 1's which is X_o, the "dummy" variable for the intercept, in the X matrix, the matrix of observations of the predetermined variables in ordinary least squares. Thus there is a linear combination among X_o and the five zero-one variables and (X'X) will be singular and it is not possible to obtain $(X'X)^{-1}$. Therefore, for estimation purposes something has to be done to alleviate the singularity problem.

There are three alternatives for handling the singularity problem.

(1) Set the sum of the parameters of the zero-one variables equal to zero. This means one of the parameters can be expressed as the sum

Table 4.1. Observations of the Five Zero-One Variables

Observation	D_1	D_2	D_3	D_4	D_5
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	1
6	1	0	0	0	0
7	0	1	0	0	0
8	0	0	1	0	0
9	0	0	0	1	0
10	0	0	0	0	1
11	1	0	0	0	0
.					
.					
.					
m	0	0	0	0	1

of the others. This is identical to the restriction on the sum of the treatment effects of a fixed treatment analysis of variance model.

(2) Eliminate the intercept from the regression model. Thus X_o would be eliminated from the X matrix.

(3) Arbitrarily choose the parameter of one of the zero-one variables to be zero which is accomplished by arbitrarily deleting one of the zero-one variables of the set. The intercept of the regression equation is the intercept pertaining to the omitted zero-one variable. It becomes the base intercept. The coefficients of the other zero-one variables indicate the differences between the base intercept and the intercept pertaining to the category of the attribute represented by a particular zero-one variable.

The third alternative is one of the most convenient and easy to incorporate. Also, it is applicable to more than one set of zero-one variables in a given regression model. If there is a more than one set of discrete variables one zero-one variable needs to be deleted from each set. If slopes are also being permitted to change, omitting one of the zero-one variables automatically omits one of the interactions of zero-one variable and continuous variable and takes care of the singularity problem arising from the interaction variables.

For purposes of estimating the parameters of the model, let us arbitrarily omit the fifth zero-one variable. Thus the first four zero-one variables, D_1, D_2, D_3 and D_4 will be included in the model and we can assume that they are the first four predetermined variables (after the "dummy" variable for the intercept) in the X matrix. The linear multiple regression model in equation form is:

$$(4.2) \quad Y = \beta_o + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \beta_5 X_5 + - - - + \beta_k X_k + \varepsilon$$

where: Y is an endogenous variable which is assumed to be a continuous quantitative variable.

D_1, D_2, D_3 and D_4 are the four zero-one predetermined variables.

X_5, X_6, X_7, --- X_k are the other predetermined variables in the model.

β's are the regression parameters.

ε is the disturbance term.

For convenience purposes, we can relabel the zero-one variables as X's, i.e., $D_1 = X_1$, $D_2 = X_2$, $D_3 = X_3$ and $D_4 = X_4$, and express the model in terms of matrices as follows:

$$(4.3) \quad Y = X\beta + \varepsilon$$

where:

Y is a (n x 1) vector of observations of the endogenous variables

X is a n x (k+1) matrix of observations of the predetermined variables including the four zero-one variables.

β is a (k+1) x 1 vector of regression parameters.

ε is a n x 1 vector of disturbances.

Furthermore, in this particular case we are considering four zero-one variables but in general the number of zero-one variables may be $h \leq k \leq n$. If $h = k$, that is all predetermined variable in the model are zero-one variables, the model is the analysis of variance model associated with a completely randomized design. When both zero-one and continuous quantitative variables are included as predetermined variables in the model, the model is a covariance model. If there is only one continuous quantitative predetermined variable the model is a simple covariance model. When there are two or more continuous quantitative predetermined variables the model is a multiple covariance model.

Now (4.3) is the representation of (4.2) in matrix notation and is the general multiple regression model discussed in Chapter 2. The equa-

tions and formuli which apply to the general regression model apply to the regression model with one set of zero-one variables used as intercept shifters.

Now let us further explore the interpretation of the coefficients of the zero-one variables and the testing of hypotheses about these particular parameters. This can be conveniently done by considering the prediction or estimated equation corresponding to (4.2) where the estimated coefficients have been obtained by least squares techniques. The estimated equation is:

$$(4.4)\quad \hat{Y} = b_o + b_1D_1 + b_2D_2 + b_3D_3 + b_4D_4 + b_5X_5 + b_6X_6 + \text{---} + b_kX_k$$

or in matrix form

$$(4.5)\quad \hat{Y} = X\,b$$

where: $D_1 = X_1$

$D_2 = X_2$

$D_3 = X_3$

$D_4 = X_4$

The b_o coefficient is the estimated intercept which pertains to the omitted category of the discrete variable. The t-test presented may be used to test hypotheses about β_o as such; specifically whether β_o is equal to or different from zero or equal to or different from some non-zero constant. b_1, b_2, b_3 and b_4 are estimates of the differences of the intercepts for categories 1, 2, 3, and 4 from the omitted category, category 5 in this case. The t-test may be used to test hypotheses about whether or not each of the respective parameters are equal to zero or equal to some non-zero constant. If β_1 is found to be significantly different from zero at some chosen probability, this indicates that the intercept of category 1 is significantly different from the intercept of the omitted category. Whether a one or two tailed alternative hypothesis is specified and whether a one or two tailed test is used depends on one's prior information. Recognize

that the estimated coefficients of the zero-one variables represent an incremental difference or change relative to a base and the base is the intercept pertaining to the omitted category. Hence, tests of hypotheses about the individual parameters of the zero-one variables provide information about whether the intercepts for each of the included categories are different from the omitted category and the estimated coefficients indicate the numerical amount by which the intercept of each of the included categories is above or below the intercept of the excluded category.

Note from (4.4) that if we consider predicted values of the endogenous variable pertaining to the fifth zero-one variable, the omitted category, that $D_1 = D_2 = D_3 = D_4 = 0$ and the appropriate prediction equation is:

$$(4.6) \quad \hat{Y} = b_o + b_5X_5 + b_6X_6 + - - - + b_kX_k$$

b_o is the relevant estimated intercept. If the predicted values of interest pertain to the first zero-one variable then $D_2 = D_3 = D_4 = 0$ and the appropriate prediction equation is:

$$(4.7) \quad \hat{Y} = (b_o + b_1) + b_5X_5 + b_6X_6 + - - - + b_kX_k$$

$(b_o + b_1)$ is the appropriate estimated intercept. For the second zero-one variable the prediction equation is:

$$(4.8) \quad \hat{Y} = (b_o + b_2) + b_5X_5 + b_6X_6 + - - - + b_kX_k$$

$(b_o + b_2)$ is the appropriate estimated intercept. For the third zero-one variable the appropriate prediction equation is:

$$(4.9) \quad \hat{Y} = (b_o + b_3) + b_5X_5 + b_6X_6 + - - - + b_kX_k$$

$(b_o + b_3)$ is the appropriate estimated intercept. And for the fourth zero-one variable the appropriate prediction equation is:

$$(4.10) \quad \hat{Y} = (b_o + b_4) + b_5X_5 + b_6X_6 + - - - b_kX_k$$

$(b_o + b_4)$ is the relevant estimated intercept.

Generally, when zero-one variables are used as intercept shifters in a regression equation one is also interested in testing hypotheses

about differences between the parameters of the zero-one variables. In other words, interest is in determining which intercepts differ significantly from each other at some chosen level of significance. Some information about differences in intercepts is provided by the tests of hypotheses about whether or not the parameters of the zero-one variables are different from zero. The hypotheses to be tested are:

HO: $\beta_j = 0 \qquad j = 0, 1, 2, 3, 4$

HA $\beta_j \neq 0$

or

HA: $\beta_j < 0$

or

HA: $\beta_j > 0$

The appropriate test is the t-test is:

$$(4.11) \qquad t = \frac{b_j - \beta_j}{s_{b_j}}$$

when $j = 0$ the hypotheses and test pertain to the intercept associated with the omitted category of the attribute. In a sense the β_o consists of two components - the intercept of the regression equation as such and the effect of the omitted category of the attribute or the fifth zero-one variable in the example considered. These two components are confounded and together considered the intercept of the omitted category. If tha null hypothesis is rejected the intercept of the omitted category is considered significantly different from zero (or some non-zero constant if the null hypothesis was specified in this manner). If the null hypothesis is not rejected one simply concludes that the sample does not provide evidence that the intercept is different from zero (or some non-zero constant).

When $j = 1$, 2, 3, or 4 the hypotheses and tests pertain to the coefficients of the zero-one variables. If the null hypothesis is rejected

in these cases the parameter is considered significantly different from zero. However, if the parameter is significantly different from zero then the intercept pertaining to that particular zero-one variable is different from the base intercept which is the one associated with the omitted category. If the null hypothesis is not rejected one concludes that the intercept pertaining to that zero-one variable is not significantly different from the base intercept, β_o. Hence, tests of the hypotheses that $\beta_j = 0$, $j = 1, 2, 3, 4$ in fact provides information about the differences between the base intercept and the intercepts associated with each of the zero-one variables.

Hypotheses about differences among the parameters of the zero-one variables may be tested, however, these are subject to the orthogonality conditions or multiple range test considerations discussed in Chapter 3. In the example, we have been considering there are four parameters of zero-one variables and hence three degrees of freedom and three orthogonal comparisons which may be submitted to statistical testing using conventional statistical tests. One set of such orthogonal contrasts are presented in Table 4.2.

Comparison	β_1	β_2	β_3	β_4	and $\sum a_{ij}$ $\sum a_{ij} a_{kj}$
I a_{1j}	+1	-1	0	0	0
II a_{2j}	0	0	+1	-1	0
a_{1j} a_{2j}	0	0	0	0	0
III a_{3j}	+1	+1	-1	-1	0
a_{1j} a_{3j}	+1	-1	0	0	0
a_{2j} a_{3j}	0	0	-1	+1	0

This set of orthogonal contrasts entails testing the difference between β_1 and β_2, the difference between β_3 and β_4, and the difference between the sum of β_1 and β_2 and the sum of β_3 and β_4. Tests of hypotheses concerning the first two comparisons could be made using the t-test presented in (3.24). The hypotheses concerning the third comparison may be tested using the t-test given in (4.11) or the F-test in (3.53).

The Newman- Kuels test presented in Chapter 3 may be used to test all possible pairs of differences among all the parameters of (4.2) or all possible pairs of differences among the parameters of the zero-one variables. If only the coefficients of the zero-one variables are considered in the example we have been considering there are six differences among estimated coefficients to be tested.

Two or More Sets of Zero-One Variables as Intercept Shifters.

For some problems it may be meaningful to incorporate two or more sets of zero-one variables to represent categories or classes of two of more attributes which are qualitative in nature or quantitative in nature for which only information about discrete categories is available. Two or more sets of zero-one variables may be incorporated into the general multiple regression model as long as the categories for each of the attributes being represented are mutually exclusive and include all possible values that each attribute can assume.

To illustrate multiple sets of zero-one variables as intercept shifters let us confine ourselves to two sets. It will be apparent that the generalization from two sets any number of sets greater than two is straight forward. As one set let us consider the five categories of an attribute discussed earlier, specified in (4.1) and finally incorporated as four zero-one variables in the regression

model (4.2). As a second set consider an attribute which consists of three categories and which would be represented by three zero-one variables F_1, F_2, and F_3. Observations for the two sets of zero-one variables are presented in Table 4.3. Note that not only does $\sum_{j=1}^{5} D_j = [1]_{nx1} = X_o$ given in (4.1) hold but also

$$(4.12) \quad \sum_{j=1}^{3} F_j = [1]_{nx1} = X_o$$

The sum of all zero-one variables in <u>each set</u> equals a column vector of 1's which is X_o, the "dummy" variable for the intercept. Prior to estimation the linear combinations between X_o and each of the sets of zero-one variables ahs to be eliminated by one of the three alternatives suggested earlier in the chapter. The same holds true for any number of sets of zero-one variables.

Let us again draw on what is probably the easiest and very convenient way of eliminating this type of linear combination problem and that is to arbitrarily choose the parameter of one the zero-one variables <u>in each set</u> to be zero. This is accomplished by arbitrarily deleting one zero-one variable <u>from each set</u>. The intercept of the equation is then the base intercept associated with the omitted zero-one variable <u>from each set.</u>

For purposes of illustrating the estimation of parameters and testing hypotheses about parameters of a regression model containing two sets of zero-one variables, let us arbitrarily omit the fifth category of the D zero-one variables and the third category of F zero-one variables. The linear multiple regression model in equation form is:

$$(4.13) \quad Y = \beta_o + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \beta_5 F_1 + \beta_6 F_2 + \beta_7 X_7 + - - + \beta_k X_k + \varepsilon$$

Table 4.3. Observations for Two Sets of Zero-One Variables With Five Categories in One Set and Three Categories in the Other Set

Observation	D_1	D_2	D_3	D_4	D_5	F_1	F_2	F_3
1	1	0	0	0	0	1	0	0
2	0	1	0	0	0	0	1	0
3	0	0	1	0	0	0	0	1
4	0	0	0	1	0	1	0	0
5	0	0	0	0	1	0	1	0
6	1	0	0	0	0	0	0	1
7	0	1	0	0	0	1	0	0
8	0	0	1	0	0	0	1	0
9	0	0	0	1	0	0	0	1
10	0	0	0	0	1	1	0	0
11	1	0	0	0	0	0	1	0
.								
.								
.								
.								
n	0	0	0	0	1	0	0	1

where:

Y is an endogenous variable which is assumed to be a continuous quantitative variable.

D_1, D_2, D_3, and D_4 are the four zero-one predetermined variables of the first set of discrete variables.

F_1 and F_2 are the two zero-one predetermined variables of the second set of discrete variables.

β's are the regression parameters.

ε is the disturbance term.

Again we can conveniently rename or relabel the zero-one variables as X's in the following way:

$$X_1 = D_1$$
$$X_2 = D_2$$
$$X_3 = D_3$$
$$X_4 = D_4$$
$$X_5 = F_1$$
$$X_6 = F_2$$

and express the model in terms of matrices as follows:

(4.14) $Y = X\beta + \varepsilon$

where: Y is a (nxl) vector of observations of the endogenous variable

X is a nx (k + 1) matrix of observations of the predetermined variables including X_o, four zero-one variables of the first set of discrete variables, and two zero-one variables of the second set of discrete variables.

Now we have the covariance model with two sets of discrete variables cast in the framework of the general multiple regression model. Also, now it may be more apparent why the first X variable in (4.13) was denoted as X_7 -- it facilitated the transition to (4.14). The equations and formuli pertaining to the general linear model discussed in Chapter 2,

apply to the multiple regression model with two sets of zero-one variables used as intercept shifters. Furthermore, at this point it should also be apparent that these equations and formuli are applicable to multiple regression models with more than two sets of zero-one variables.

The interpretation of estimated coefficients of the zero-one variables and the testing of hypotheses about these parameters when there are two or more sets of zero-one variables are similar as in the case where there is only one set of zero-one variables. Let us focus on the estimated equation corresponding to (4.13) which is:

$$(4.15)\quad \hat{Y} = b_0 + b_1D_1 + b_2D_2 + b_3D_3 + b_4D_4 + b_5F_1 + b_6F_2 + b_7X_7 + \text{---} + b_kX_k$$

The b_0 is the estimated intercept pertaining to the omitted categories of the two sets of discrete variables. In this case omitted categories are the fifth of the D set and the third of the F set. In a sense this intercept embodies three confounded components - some general intercept for the equation, the effect of the omitted D variable, D_5 and the omitted F variable, F_3. The t-test presented in (4.11) may be used to test hypotheses about β_0, specifically whether β_0 is equal to zero or equal to some non-zero constant or whether it is significantly different at some chosen level of significance. The test of hypotheses about parameters of the zero-one variables of both sets procedes along the same lines discussed for the case where there was only one set of zero-one variables. The estimated coefficients of the zero-one variables indicate the numerical amount by which the intercept of the included categories of the two sets of discrete variables differ from the base intercept which pertains to an omitted category from each set.

With reference to (4.15) if we consider predicted values of the endogenous variable pertaining to the fifth zero-one variable of the first set of discrete variables and the third zero-one variable of the second set of discrete variables the appropriate prediction equation is:

(4.16) $\hat{Y} = b_0 + b_7X_7 + b_8X_8 + - - - + b_kX_k$

Remember for this situation $D_1 = D_2 = D_3 = D_4 = F_1 = F_2 = 0$. b_0 is the relevant estimated intercept. If the predicted value pertains to the first zero-one variable of the first set of discrete variables and the first zero-one variable of the second set of discrete variables the appropriate prediction equation is:

(4.17) $\hat{Y} = (b_0 + b_1 + b_5) + b_7X_7 + b_8X_8 + - - - + b_kX_k$

$(b_0 + b_1 + b_5)$ is the appropriate estimated intercept. The estimated equation for predicted values of the endogenous variable pertaining to the included catefories of the two attributes represented by the two sets of discrete variables will contain an intercept consisting of three components, b_0, one of the four estimated coefficients, b_1, b_2, b_3, or b_4, and either b_5 or b_6.

Hypotheses about differences among the parameters of each set of zero-one variables may be tested, however, they are subject to the orthogonality conditions or multiple range test considerations discussed in Chapter 3. Orthogonal comparisons of the parameters for each set of zero-one variables or the Newman-Kuels test applied to the estimated coefficients of each set of zero-one variables, estimated coefficients of each set of zero-one variables, estimated coefficients of all sets of zero-one variables, or all estimated coefficients of the regression equation as discussed in Chapter 3 may be made.

One Set of Zero-One Variables as Slope Shifters

There may be reasons in some problems to allow for differences in slopes of one or more of the continuous predetermined variables for various categories of some qualitative variable or categorized quantitative variable. This may be incorporated into a multiple regression model by including certain interaction variables which are generated by interacting a set of zero-one variables with the selected quantitative predetermined variables of interest. Allowing for differences in slopes should not be done indiscreminately. There should be a sound basis, theoretical or other, for considering differences in slopes of continuous predetermined variables with respect to categories of qualitative variables or categorized quantitative variables. Otherwise, the final statistical results may just be nonsense. In some cases they may be quite meaningful. For instance, the qualitative variable may represent regions. One may be investigating the same type of model for all the regions and there may be a reason to believe and hypothesize that the change in the endogenous variable associated with some given incremental change in a particular predetermined variable may be different for the different regions. An alternative to estimating the parameters of the regression model separately for each of the regions is to estimate a model applicable to all regions using the data for all regions pooled together and allowing the slope or slopes of selected predetermined variables to be different among regions. This may be done in conjunction with a common intercept for all regions or the intercept may be allowed to vary by use of zero-one variables as intercept shifters.

Let us consider the case in which there is a common intercept (intercepts are not allowed to shift by the use of zero-one variables) but the

slopes of the continuous predetermined variables are allowed to change. To keep the illustration manageable, let us consider an attribute consisting of five categories and let us select two continuous predetermined variables whose slopes we will allow to change. To allow slopes to change among the categories of the attribute requires generating new variables from zero-one variables representing the categories of the attribute and the quantitative predetermined variables whose coefficients we are allowing to vary. The new variables are generated by constructing the first order interaction between each of the zero-one variables and the selected quantitative predetermined variables. In other words the new variables are generated by forming cross products between each of the zero-one variables representing the categories of the attribute and the selected continuous predetermined variables whose coefficients are being allowed to vary among categories.

To illustrate the use of zero-one variables for allowing slope changes let us consider the five D variables whose observations are presented in Table 4.1 and which are further defined in (4.1). Also, let us restrict the slope changes to the first two quantitative predetermined variables. The observations of the two predetermined variables and the generated interaction variables are presented in Table 4.4. Note that:

$$(4.18) \qquad (X_1D_1) + (X_1D_2) + (X_1D_3) + (X_1D_4) + (X_1D_5) = X_1$$

and

$$(4.19) \qquad (X_2D_1) + (X_2D_2) + (X_2D_3) + (X_2D_4) + (X_2D_5) = X_2$$

Thus the sum of the interactive (crossproduct) variables and each of the quantitative predetermined variables evolve in a linear combination. Hence, all categories of the attribute must not be represented by zero-one

Table 4.4. Observations of First Two Quantitative Predetermined Variables and Generated Interaction Variables

Observation	X_1	X_2	(X_1D_1)	(X_1D_2)	(X_1D_3)	(X_1D_4)	(X_1D_5)	(X_2D_1)	(X_2D_2)	(X_2D_3)	(X_2D_4)	(X_2D_5)
1	X_{11}	X_{12}	X_{11}	0	0	0	0	X_{12}	0	0	0	0
2	X_{21}	X_{22}	0	X_{21}	0	0	0	0	X_{22}	0	0	0
3	X_{31}	X_{32}	0	0	X_{31}	0	0	0	0	X_{32}	0	0
4	X_{41}	X_{42}	0	0	0	X_{41}	0	0	0	0	X_{42}	0
5	X_{51}	X_{52}	0	0	0	0	X_{51}	0	0	0	0	X_{52}
6	X_{61}	X_{62}	X_{61}	0	0	0	0	X_{62}	0	0	0	0
7	X_{71}	X_{72}	0	X_{71}	0	0	0	0	X_{72}	0	0	0
8	X_{81}	X_{82}	0	0	X_{81}	0	0	0	0	X_{82}	0	0
9	X_{91}	X_{92}	0	0	0	X_{91}	0	0	0	0	X_{92}	0
10	$X_{10,1}$	$X_{10,2}$	0	0	0	0	$X_{10,1}$	0	0	0	0	$X_{10,2}$
11	$X_{11,1}$	$X_{11,2}$	$X_{11,1}$	0	0	0	0	$X_{11,2}$	0	0	0	0
.	.	.										
.	.	.										
.	.	.										
.	.	.										
n	$X_{n,1}$	$X_{n,2}$	0	0	0	0	$X_{n,1}$	0	0	0	0	$X_{n,2}$

variables. The simpliest way to alleviate this problem is to omit one of the zero-one variables representing one of the categories of the attribute. The choice is arbitrary. By omitting one of the zero-one variables one interaction variable pertaining to X_1 and one pertaining to X_2 are automatically deleted and the linear dependency is dissapated. Then one can procede with the estimation of parameters of the multiple regression model using least squares methods.

To illustrate the estimation of parameters of such a model, let us omit the fifth category of the attribute which is represented by D_5. Automatically (X_1D_5) and (X_2D_5) are deleted. Now (4.18) and (4.19) no longer hold. The linear multiple regression model in equation form is:

$$(4.20) \quad Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3(X_1D_1) + \beta_4(X_1D_2) + \beta_5(X_1D_3)$$

$$+ \beta_6(X_1D_4) + \beta_7(X_2D_1) + \beta_8(X_2D_2) + \beta_9(X_2D_3) + \beta_{10}(X_2D_4)$$

$$+ \beta_{11}X_{11} + - - - + \beta_kX_k + \varepsilon$$

where:

Y is an endogenous variable which is assumed to be a continuous quantitative variable.

(X_1D_1), (X_1D_2) - - - (X_2D_3) (X_2D_4) are the generated interaction variables.

X_1, X_2, X_{11}, X_{12}, - - - X_k are the quantitative predetermined variables in the model.

β's are the regression parameters.

ε is the distumbance term.

Again for convenience purposes, we can relabel the interaction variables as X's, i.e., X_3 through X_{10} and the model may be expressed in the familiar matrix form: $Y = \beta X + \epsilon$. Estimation of psrameters and testing of hypotheses about the parameters msy be performed following conventional procedures fo the least squares framework.

The interpretation of the estimated coefficients merits a brief discussion. To do this let us consider the estimated equation corresponding to (4.20) which is:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3(X_1D_1) + b_4(X_1D_2) + b_5(X_1D_3) + b_6(X_1D_4) + b_7(X_2D_1) + b_8(X_2D_2) + b_9(X_2D_3) + b_{10}(X_2D_4) + b_{11}X_{11} + - - - + b_kX_k. \tag{4.21}$$

Now let us consider a predicted value of Y which pertains to the omitted category of the attribute. $D_1 = D_2 = D_3 = D_4 = 0$ and relevant prediction equation is:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11} + - - - + b_kX_k \tag{4.22}$$

Note that b_1 and b_2 are the slope coefficients of X_1 and X_2 respectively sssociated with omitted category of attribute which could be represented by a zero-one variables. They are base slopes. b_1 indicates the amount of chsnge in Y associated with a one unit change in X_1 subject to the effects of the omitted category of the attribute. b_2 indicates the amount of change in Y associated with a one unit change in X_2 subject to the effects of the omitted category of the attribute. b_0 is simply the estimated intercept of the equation. b_{11} through b_k are estimated regression coefficients of X_{11} through X_k and interpreted in the usual manner.

Now let us consider predicted value of the endogenous varisble associated with the first category of the qualitative or categorized quantitative

attribute. $D_2 = D_3 = D_4 = 0$ and the appropriate estimated equation is:

$$(4.23)\quad \hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3(X_1D_1) + b_7(X_2D_1) + b_{11}X_{11} + \text{---} + b_kX_k$$

But $D_1 = 1$ so

$$(4.24)\quad \hat{Y} = b_0 + (b_1+b_3)\,X_1 + (b_2+b_7)\,X_2 + b_{11}X_{11} + - - - + b_kX_k$$

The amount of change in Y associated with a one unit change in X_1 subject to the effects of D_1 is $(b_1 + b_3)$, the slope coefficient of X_1 for the first discrete variable. Similarly, the slope coefficient of X_2 for D_1 is $(b_2 + b_7)$. b_3 and b_7 are estimates of the differences in slopes from the base slopes b_1 and b_2 associated with the first discrete variable. The slope coefficients for the other discrete variables are obtained and interpreted in a similar manner.

Tests of hypotheses regarding whether or not the parameters of the interaction variables are equal to zero or some non-zero constant may be performed using a t-test. These tests provide information about whether or not the slopes associated with the discrete variables are significantly different at some chosen level of significance from the base slopes. Tests of hypotheses about differences among pairs of other linear combinations of the parameters are subject to the orthogonality conditions or multiple range test considerations discussed in Chapter 3.

Two or More Sets of Zero-One Variables as Slope Shifters

There is nothing that precludes having more than one set of discrete variables as slope shifters. If they are meaningful and helpful they may be used. Let us briefly explore the case of two sets using the five categories represented by the D variables just discussed and the three categories of a second set which were introduced in Table 4.3 and denoted as F variables.

Again in order to keep the model manageable, let us restrict the allowance of slope differences to X_1 and X_2. However, recognize that the slopes of all continuous predetermined variables could be allowed to change with respect to categories of qualitative or categorized quantitative variables. The interaction variables representing cross products of the D variables with X_1 and X_2 are already presented in Table 4.3. Similar variables have to be generated for the F variables and X_1 and X_2. In terms of the observations presented in Table 4.4. Similar variables have to be generated for the F variables and X_1 and X_2. In terms of the observations presented in Table 4.4 there would be 6 additional columns consisting of (X_1F_1), (X_1F_2), (X_1F_3), (X_2F_1), (X_2F_2) and (X_2F_3). However, similar to (4.18) and (4.19)

$$(4.25) \qquad (X_1F_1) + (X_1F_2) + (X_1F_3) = X_1$$

and

$$(4.26) \qquad (X_2F_1) + (X_2F_2) + (X_2F_3) = X_2$$

The sum of the interaction variables in this case F and X, and the two continuous predetermined variables results in a linear combination. Hence, again one of the categories of the F variable needs to be omitted. Let us arbitrarily omit F_3. This eliminate (X_1F_3) and (X_2F_3). The linear multiple regresaion model in equation form is:

$$\begin{aligned}(4.27) \qquad Y = {} & \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3(X_1D_1) + \beta_4(X_1D_2) + \beta_5(X_1D_3) + \beta_6(X_1D_4) \\ & + \beta_7(X_2D_1) + \beta_8(X_2D_2) + \beta_9(X_2D_3) + \beta_{10}(X_2D_4) + \beta_{11}(X_1F_1) \\ & + \beta_{12}(X_1F_2) + \beta_{13}(X_2F_1) + \beta_{14}(X_2F_2) + \beta_{15}X_{15} + \beta_{16}X_{16} + - - - \\ & + \beta_kX_k + \varepsilon\end{aligned}$$

where Y is the endogenous variable which is assumed to be a continuous variable.

(X_1D_1), (X_1D_2) - - - (X_2D_3) (X_2D_4) are the generated interaction variables pertaining to the D discrete variables.

(X_1F_1), (X_1F_2), (X_2F_1) and (X_2F_2) are the generated interaction variables pertaining the the F discrete variables.

X_1, X_2, X_{15}, X_{16} - - - X_k are the continuous predetermined variables in the model.

β's are the regression parameters.

ε is the disturbance term

Now we may relabel the interaction variables as X's and express the model form as $Y = X\beta + \varepsilon$ which is the general multiple regression model. Estimation may be performed according to least squares methods and testing of hypotheses about the parameters may be performed using conventional tests as long as orthogonality conditions are met.

Interpretation of coefficients is similar to the case where only one set of zero-one variables was used to shift slopes. Consider the estimated equation corresponding to (4.27) which is:

$$(4.28)\quad \hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3(X_1D_1) + b_4(X_1D_2) + b_5(X_1D_3) + b_6(X_1D_4) + b_7(X_2D_1) + b_8(X_2D_2) + b_9(X_2D_3) + b_{10}(X_2D_4) + b_{11}(X_1F_1) + b_{12}(X_1F_2) + b_{13}(X_2F_1) + b_{14}(X_2F_2) + b_{15}X_{15} + - - - + b_kX_k$$

when $D_1 = D_2 = D_3 = D_4 = F_1 = F_2 = 0$ the predicted values of the endogenous variable pertain to the omitted categories of D and F which are D_5 and F_3. The relevant prediction equation is:

$$(4.29)\quad \hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_{15}X_{15} + - - - + b_kX_k$$

b_0 is the estimated intercept of the equation and is common to all of the

discrete variables. b_1 and b_2 are the slope coefficients of X_1 and X_2 respectively associated with the omitted D and F variables. They are base slopes. b_1 indicates the amount of change in Y associated with a one unit change in X_1 subject to the effects of D_5 and F_3. b_2 is interpreted similarly for X_2.

Now let us consider the predicted values of Y associated with the first D variable and the first Y variable. $D_2 = D_3 = D_4 = F_2 = 0$ and the relevant estimated equation is:

$$(4.30)\quad \hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3(X_1D_1) + b_7(X_2D_1) + b_{11}(X_1F_1) + b_{13}(X_2F_1) + b_{15}X_{15} + - - - + b_kX_k$$

But $D_1 = F_1 = 1$ so

$$(4.31)\quad \hat{Y} = b_0 + (b_1 + b_3 + b_{11})\,X_1 + (b_2 + b_7 + b_{15})\,X_2 + b_{15}X_{15} + - - - + b_kX_k$$

The amount of change in Y associated with a unit change in X_1 subject to the effects of D_1 and F_1 is $(b_1 + b_3 + b_{11})$, the slope coefficient. Similarly, the slope coefficient of X_2 is $(b_2 + b_7 + b_{13})$. b_3 and b_7 are estimates of differences in slopes from the base slopes associated with D_1 and b_{11} and b_{13} are differences in slopes from the base slopes associated with F_1. The slopes coefficients for the other discrete variables are obtained and interpreted in a similar manner.

Tests of hypotheses about the parameters of the interaction varisbles and other predetermined variables in the model are the same as those discussed for a single set of zero-one variables used to shift slopes.

One Set of Zero-One Variables as Intercept and Slope Shifters

For problems in which it is meaningful to all for differences in slopes it may be meaningful to concurrently allow for differences in intercept. Let us consider the case where a qualitative variable or a categorized quantitative variable with five catefories is used to allow both intercepts and slopes of two predetermined variables to differ. This is basically incorporating the information from Table 4.1 and Table 4.4 in a multiple regression model. Let us arbitrarily omit D_5. The regression equation of interest is:

$$(4.32)\quad Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \beta_5 X_1 + \beta_6 X_2 + \beta_7 (X_1 D_1)$$
$$+ \beta_8 (X_1 D_2) + \beta_9 (X_1 D_3) + \beta_{10} (X_1 D_4) + \beta_{11} (X_2 D_1) + \beta_{12} (X_2 D_2)$$
$$+ \beta_{13} (X_2 D_3) + \beta_{14} (X_2 D_4) + \beta_{15} X_{15} + - - - + \beta_k X_k + \varepsilon$$

where: Y is the endogenous variable which is assumed to be a continuous variable.

D_1, D_2, D_3 and D_4 are four zero-one predetermined variables.

$(X_1 D_1)$, $(X_1 D_2)$, - - - $(X_2 D_3)$, $(X_2 D_4)$ are the generated interaction variables.

X_1, X_2, X_{15} - - - X_k are the continuous predetermined variables in the model.

β's are the regression parameters.

ε is the distrubance term.

Again by relabeling the interaction variables as X's i.e., X_3 through X_{14} the model may be expressed in matrix form as $Y = X\beta + \varepsilon$.

The equation with parameter estimates obtained by least squares methods is:

$$(4.33)\quad \hat{Y} = b_0 + b_1D_1 + b_2D_2 + b_3D_3 + b_4D_4 + b_5X_1 + b_6X_1 + b_7(X_1D_1)$$
$$+ b_8(X_1D_2) + b_9(X_1D_3) + b_{10}(X_1D_4) + b_{11}(X_2D_1) + b_{12}(X_2D_2$$
$$+ b_{13}(X_2D_3) + b_{14}(X_2D_4) + b_{15}X_{15} + - - - + b_kX_k$$

The interpretation of estimated coefficients and tests of hypotheses about the paraters are the same as those discussed earlier in this chapter.

Two or More Sets of Zero-One Variables As Intercept and Slope Shifters

Two or more sets of zero-one variables may be used to allow for differences in both intercepts and slopes. This simply entails bringing the information concerning two sets of intercepts shifters and two sets of slope shifters together into one multiple regression equation. The generalization to more than two sets simply entails following the same procedures that pertain to the two set case and expanding to the case with more than two sets. Details are not presented here since they should be apparent from the details of the component parts (two sets of intercept shifters and two sets of slope shifters) presented earlier.

Concluding Remarks About Zero-One Variables As Intercept and Slope Shifters

The mechanics of allowing for intercept and slope difference with respect to categories of qualitative or categorized quantitative variables is quite simple. However, these techniques should be used with caution. The analyst needs to clearly think through the justification for differences in intercepts and slopes. If the parameters represent some economic structure then the question of why there should be such differences or changes in structure has to be answered. If there is a

theoretical or other justification of such differences or changes then the different intercepts and slopes may be meaningful. Otherwise even though significant differences may be identified statistically from an economic viewpoint they may be nonsense.

Although only two sets of intercept shifters were illustrated there may be any number as long as adequate degrees of freedom are available. Also, only five and three categories of the qualitative or categorized quantitative variables were considered. The number of categories can range from two which can be represented by including one zero-one variable in the model to any number provided adequate degrees of freedom exist.

The illustrations concerning shifts or differences in slopes confined the slope changes to two continuous predetermined variables. This was done strictly to keep the illustrations manageable. The slopes of all continuous predetermined variables may be allowed to shift or take on different values for different zero-one variables. This may be done for a single set of discrete variables or a multiple set.

The use of zero-one variables particularly when slopes are allowed to vary requires a large number of degrees of freedom. In fact the degrees of freedom are one factor which may limit the number of predetermined variables whose slopes will be allowed to vary. Even if degrees of freedom aren't a limiting factor the size of the equation which arises out of allowing slopes of a large number of predetermined variables to vary may create difficulties of analysis and interpretation. Also, some extremely high intercorrelations may arise from large numbers of interaction variables required to allow for differences in slopes of a large number of predetermined variables. Of

course, in terms of the overall analysis the use of zero-one variables facilitates pooling of data for estimation and testing purposes so that there may be a gain in degrees of freedom as compared to estimating separate relationships for each set of data.

It might be worth pointing out that allowing intercepts and the slopes of all predetermined variables to shift is equivalent to fitting separate regression equations to the data for each category of the qualitative or categorized quantitative variables.

Finally, throughout the discussion there was considerable stress on the arbitrary choice of omitting a category of the qualitative or categorized quantitative variable. This simply means the analyst is at liberty to choose whichever ones he pleases. Under certain conditions some choices amy be considerably better than others. This is particularly true if one has some prior information about the magnitude of the shifts in intercepts or slopes. If prior information is available then from the viewpoint of efficiency in testing and drawing inferences, it is to one's advantage not to omit an extreme category. By omitting a sort of "in the middle" category one may be able to make inferences about differences between intercepts or slopes without testing hypotheses about the differences between pairs of parameters. For example, if one concludes that a parameter is greater than zero while for another parameter the conclusion is that it is less than zero than one can conclude that the two parameters are significantly different from each other without directly testing such hypotheses. However, if an extreme category had been omitted then the parameters may have been found to be significantly greater than or significantly less than zero with little information about the difference between the parameters.

Matrix Representation of Zero-One Variables As Intercept and Slope Shifters

Let us attempt to summarize the use of zero-one variables as intercept and/or slope shifters in a compact form. This may be done by utilizing partitioned matrices. Let the multiple regression model under consideration be:

(4.34) $Y = Z\beta + \varepsilon$

When none of the predetermined variables are zero-one variables then

(4.35) $Z = X$

and

(4.36) $Y = Z\beta + \varepsilon = X\beta + \varepsilon$

We simply have the multiple regression model expressed in terms of matrices.

Now let us consider one set of zero-one variables to be used for allowing for differences in intercepts. Let us assume the linear dependency problem has been circumvented by arbitrary omitting one category of the attribute by deleting one of the zero-one variables. Denoting the matrix of observations of the included zero-one variables as D, the D is a sub-matrix of the matrix of observations of all predetermined variables and may be expressed as a partition of Z as follows:

(4.37) $Z = (D:X)$

where: D is the matrix of observations of the included zero-one variables.

X is the matrix of observations of the predetermined variables that are not zero-one variables.

The X_0 variable consisting of a vector of one's is included in X and for convenience may be considered the last X variable. The regression model may be written as:

(4.38) $$Y = (D : X) \left[\begin{array}{c} \beta_1 \\ \hline \beta_2 \end{array}\right] + \varepsilon$$

where: Y is the vector of observations of the endogenous variable

$(D : X) = Z$ is defined in (4.37)

$\begin{bmatrix} \beta_1 \\ \hline \beta_2 \end{bmatrix}$ is the partitioned vector of regression parameters with β_1 referring to the parameters of the zero-one variables and β_2 to the parameters of the X variables and β_2 includes the base intercept.

Following least squares procedures the estimators of the regression parameters are:

$$(4.39) \qquad \begin{bmatrix} b_1 \\ \hline b_2 \end{bmatrix} = (Z'Z)^{-1} Z'Y = [(D : X)' (D : X)]^{-1} (D : X)' Y$$

Statistical quantities of interest such as sums of squares errors, sums of squares regression, R^2, residual variance, etc. may be obtained by procedures discussed in Chapter 2. The estimated variance-covariance matrix of the estimators is:

$$(4.40) \qquad V(b) = s^2 [(D : X)' (D : X)]^{-1}$$

which provides the variances and standard errors needed to test hypotheses about the individual parameters and for making interval estimates for the individual parameters and predicted values of the endogenous variable.

The generalization to any number of sets of zero-one variables used as intercept shifters is direct and straightforward. Let D be partitioned into f partitions such as

$$(4.41) \qquad D = (D_1 \vdots D_2 \vdots - - - \vdots D_f)$$

where each D represents a set of zero-one variables. There may be any number of zero-one variables in each set. The multiple regression model for this case is:

$$(4.42) \quad Y = (D_1 : D_2 : - - - - : D_f : X) \begin{bmatrix} \beta_1 \\ \hline \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_f \\ \hline \beta_{f+1} \end{bmatrix} + \quad \varepsilon$$

where β_1, β_2 - - - β_f are the sets of parameters of the zero-one variables and β_{f+1} are the parameters of X. The least squares estimators are:

$$(4.43) \quad \begin{bmatrix} b_1 \\ \hline b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_f \\ \hline b_{f+1} \end{bmatrix} = [(D_1 : D_2 : - - - - : D_f : X)' (D_1 : D_2 : - - - - : D_f : X)]^{-1} (D_1 : D_2 : - - - - : D_f : X)' Y$$

Other statistical quantities of interest may be obtained as discussed above. The variance-covariance matrix of the estimators given in (4.43) is:

$$(4.44) \quad V(b) = s^2 [(D_1 : D_2 : - - - - : D_f : X)' (D_1 : D_2 : - - - - : D_f : X)]^{-1}$$

In order to incorporate zero-one variables to allow for slope changes of some of the predetermined variables partition the X matrix as follows:

$$(4.45) \quad X = (X_1 : X_2)$$

where: X_1 is the set of predetermined variables whose slopes are going to be allowed to change.

X_2 is the set of predetermined variables whose slopes are not going to be allowed to change.

X_1 consists of one or more predetermined variables. If the slopes of all predetermined variables are going to be allowed to shift then the X matrix is not partitioned. Considering only one set of zero-one variables as slope shifters then:

(4.46) $\quad Z = (D'X_1 \vdots X_1 \vdots X_2)$

and

(4.47) $\quad Y = (D'X_1 \vdots X_1 \vdots X_2) \begin{bmatrix} \beta_1 \\ \hline \beta_2 \\ \hline \beta_3 \end{bmatrix} + \varepsilon$

where: β_1 are the parameters of the slope shifting variables, $D'X_1$, the generated interaction variables.

β_2 are the base slopes of the predetermined variables whose slopes are being allowed to change.

β_3 are the parameters of the predetermined variables whose slopes are not being allowed to change.

The estimators of the parameters of (4.47) are:

(4.48) $$\begin{bmatrix} b_1 \\ \hline b_2 \\ \hline b_3 \end{bmatrix} = [(D'X_1 \vdots X_1 \vdots X_2)' (D'X_1 \vdots X_1 \vdots X_2)]^{-1} (D'X_1 \vdots X_1 \vdots X_2)Y$$

The variance-covariance matrix of the estimators of (4.48) is:

(4.49) $\quad V(b) = s^2 [(D'X_1 \vdots X_1 \vdots X_2)' (D'X_1 \vdots X_1 \vdots X_2)]^{-1}$

If more than one set of zero-one variables are going to be used to allow for slope differences then the matrix of predetermined variables is:

(4.50) $\quad Z = (D_1'X_1 \vdots D_2X_1 \vdots - - - D_f'X_1 \vdots X_1 \vdots X_2)$

and the model is:

$$(4.51) \qquad Y = (D_1'X_1 \vdots D_2'X_1 \vdots \cdots \vdots D_f'X_1 \vdots X_1 \vdots X_2) \begin{bmatrix} \beta_1 \\ \hline \beta_2 \\ \hline \vdots \\ \beta_f \\ \hline \beta_{f+1} \\ \beta_{f+2} \end{bmatrix} + \varepsilon$$

where $\beta_1 \cdots \beta_f$ are the sets of parameters of the f sets of slope shifting variables

β_{f+1} is the set of base slopes of the predetermined variables whose slopes are being allowed to change.

β_{f+2} is the set of parameters of the predetermined variables whose slopes are not being allowed to change

The estimators of the parameters of (4.51) are:

$$(4.52) \qquad \begin{bmatrix} b_1 \\ \hline b_2 \\ \hline \vdots \\ \hline b_f \\ \hline b_{f+1} \\ b_{f+2} \end{bmatrix} = (Z'Z)^{-1} Z'Y$$

where Z is given in (4.50). The variance-covariance matrix of the estimators in (4.52) is:

$$(4.53) \qquad V(b) = s^2 (Z'Z)^{-1}$$

Now to consider the general case of allowing for both intercept and slope changes for some f sets of zero-one variables, then:

$$(4.54) \quad Z = (D_1 \vdots D_2 \vdots - - - \vdots D_f \vdots D_1'X_1 \vdots D_2'X_1 \vdots - - - \vdots D_f'X_1 \vdots X_1 \vdots X_2)$$

and the model is:

$$(4.55) \quad Y = (D_1 \vdots D_2 \vdots - - - \; D_f \vdots D_1'X_1 \vdots D_2'X_1 \vdots - - - \vdots D_f'X_1 \vdots X_1 \vdots X_2) \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \beta_f \\ \beta_{f+1} \\ \cdot \\ \beta_{2f} \\ \beta_{2f+1} \\ \beta_{2f+2} \end{bmatrix} + \varepsilon$$

where: $\beta_1 - - - \beta_f$ are the sets of parameters of the zero-one variables which are the intercept shifting parameters.

$\beta_{f+1} - - - \beta_{2f}$ are the sets of parameters of the f sets of slope shifting variables

β_{2f+1} is the set of base slopes of the predetermined variables whose slopes are being allowed to change.

β_{2f+2} is the set of parameters of the predetermined variables whose slopes are not being allowed to change.

The estimators of the parameters of (4.55) are:

$$(4.56) \qquad \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_f \\ b_{f+1} \\ \vdots \\ b_{2f} \\ b_{2f+1} \\ b_{2f+2} \end{bmatrix} = (Z'Z)^{-1} Z'Y$$

where Z is defined in (4.50). The variance-covariance matrix of the estimators in (4.56) is:

$$(4.57) \qquad V(b) = s^2 (Z'Z)^{-1}$$

Other statistical quantities, tests of hypotheses, and interval estimates are obtained by the procedures of least squares discussed previously.

Zero-One Dependent Variable

Thus far we have discussed the use of zero-one variables as predetermined variables but the endogenous variable may also be a zero-one variable. In cases where the endogenous variable represents the occurence or non-occurrence of an event, the endogenous variable will take on values of one or zero and is a zero-one variable. For example, a person or spending unit owns a car or does not, has a mortgage on a home or does not, belongs to an organization or does not, ranks a certain objective as number one or does not, etc. For such problems a multiple regression model may be formulated

in which the zero-one endogenous variable is regressed on several predetermined. The predetermined variables may be continuous variables or some or all may be discrete variables.

Least squares methods apply to multiple regression models in which the endogenous variable is a zero-one variable representing two classes, the occurrence or non-occurrence of an event. In these models the estimated or calculated value of the endogenous variable is interpreted as a conditional probability. It is the conditional probability of the occurrence of the endogenous variable given a set of values of the X, the predetermined variables.

A problem arises in applying classical least squares methods to a regression model where the endogenous variable takes on values of one or zero. The problem is that the assumption of homoscedastic disturbances is untendable. Note that

(4.58) $Y = X\beta + \varepsilon$

and for any $t^{\underline{th}}$ observation

(4.59) $Y_t = X_t'\beta + \varepsilon_t$

where X_t' is a particular row of X's.

and

(4.60) $\varepsilon_t = Y_t - X_t'\beta$

but Y_t is either zero or one so

(4.61) $\varepsilon_t = 1 - X_t'\beta$ when $Y_t = 1$

or

(4.62) $\varepsilon_t = -X_t'\beta$ when $Y_t = 0$

Since ε_t takes on the same distribution as Y_t, the assumption of normality cannot be fulfilled since the distribution of ε is binomial.

Now denoting the probability that $Y_t = 1$ as P_1 and the probability that $Y_t = 0$ as P_2 we have

(4.63) $$E(\varepsilon_t) = (1 - X_t'\beta)\, P_1 + (-X_t'\beta)\, P_2$$

where

(4.64) $$P_1 + P_2 = 1$$

and

(4.65) $$P_1 = 1 - P_2$$

Now under the assumption that $E(\varepsilon_t) = 0$ then (4.63) with (4.65) substituted for P is:

(4.66) $$(1 - X_t'\beta)\,(1 - P_2) + (-X_t'\beta) = 0$$

$$1 - X_t'\beta - P_2 = 0$$

$$P_2 = 1 - X_t'\beta$$

Also

(4.67) $$P(\varepsilon_t = -X_t'\beta) = 1 - P_2 = 1 - (1 - X_t'\beta) = X_t'\beta$$

Therefore, if $E(\varepsilon_t) = 0$ the distribution of ε_t, given x_t is:

(4.68)

ε_t	Probability of ε_t
$-X_t'\beta$	$1 - X_t'\beta$
$1 - X_t'\beta$	$X_t'\beta$

and the variance of ε_t is

(4.69) $$\text{Var}(\varepsilon_t) = E(\varepsilon_t^2) = E\,[\varepsilon_{t1}^2\, f(\varepsilon_{t1}^2) + \varepsilon_{t0}^2\, f(\varepsilon_{t0})]$$

$$= (-X_t'\beta)^2\,(1 - X_t'\beta) + (1 - X_t'\beta)^2\,(X_t'\beta)$$

$$= (X_t'\beta)^2 - (X_t'\beta)^3 + [1 - 2X_t'\beta + (X_t'\beta)^2]\,(X_t'\beta)$$

$$= (X_t'\beta)^2 - (X_t'\beta)^3 + X_t'\beta - 2\,(X_t'\beta)^2 + (X_t'\beta)^3$$

$$= X_t'\beta - (X_t'\beta)^2 = X_t'\beta\,(1 - X_t'\beta) = E[Y_t\,(1 - Y_t)]$$

The variance-covariance matrix of the disturbances is:

$$(4.70) \qquad E(\varepsilon\varepsilon') = \begin{bmatrix} X_1'\beta(1 - X_1'\beta) & 0 & 0 & \cdots & 0 \\ 0 & X_2'\beta(1-X_2'\beta) & & & 0 \\ 0 & & & & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & & X_n'\beta(1 - X_n'\beta) \end{bmatrix}$$

Note that the values of the variances and covariances of the disturbances depend on the values of X_t' and vary with the values of the observations of the predetermined variables. Thus, the variances are heteroscadastic depending on the values of X_t and thus the assumption of homogeneous variances cannot be fulfilled.

To obtain efficient linear estimators, procedures which take into consideration heterogeneous variances need to be used.[1/] Generalized least squares procedures, which we will discuss later could be used. Also, for some problems the unbiased property may not be so crucial nor mandatory.

Discriminate Analysis

Regression analysis entailing a zero-one endogenous variable where the endogenous variable can take on values from two classes, occurrence or non-occurrence of an event, may be used as a <u>discriminate analysis</u>. Discriminate analysis concerns itself with classifying individuals, firms, or other types of units into classes according to information about various variables and characteristics of the individuals, firms or units, For example if we have information on family size income, composition, place of residence, and other such characteristics we may be interested in predicting whether a family will or will not own an automobile. Or if we have detailed information

---

[1/] See J. Johnston, <u>Econometric Methods</u>, Second Edition, McGraw-Hill Book Co., Inc. New York, 1972, pp. 182-186 and p. 208-221 and Arthur Goldberger, <u>Econometric Theory</u>, Wiley, 1964, p. 248-51 and Jan Kmenta, <u>Elements of Econometrics</u>, 1971, pp. 409-30, and p.p. 499-508.

about the characteristics of the farmer, his family, and his operation we may be interested in predicting whether he will or will not participate in a given farm program. These two examples are illustrations of discriminate problems.

Now to symbollically illustrate a simple discriminate analysis consider an endogenous variable, Y, which takes on either a value of one or zero, that is two classes or categories define all values which Y can take on. And suppose we have some K predetermined variables X_k. The endogenous variable could represent whether families or spending units own or do not own an automobile. The predetermined variables could be such things as family size, income, family composition, place of residence, etc. The discriminate function of interest is a regression model of the form:

$$Y = X\beta + \varepsilon \tag{4.71}$$

The estimated regression equation is:

$$\hat{Y} = X b \tag{4.72}$$

where b is the vector of estimators of β in (4.71) obtained by least squares procedures. The discriminate analysis entails classifying the endogenous variable into either category one or two for given values of the set of predetermined variables, X. The estimated discriminate function for any $t^{\underline{th}}$ value of Y is:

$$\hat{Y}_t = b_0 + b_1 X_{t1} + b_2 X_{t2} + - - - + b_k X_{tk} \tag{4.73}$$

Any observation for which

$$\hat{Y}_t \geq b_0 + 1/2 \sum_{i=1}^{k} b_i \ (\bar{X}_{i1} + \bar{X}_{i2}) \tag{4.74}$$

$$\geq b_0 + 1/2 \ [b_1 \ (\bar{X}_{11} + \bar{X}_{12}) + b_2(\bar{X}_{21} + \bar{X}_{22}) + --- + b_k(\bar{X}_{k1} + \bar{X}_{k2})]$$

is classified into category one (for example own an automobile) and any observation for which

$$(4.75) \quad \hat{Y}_t < b_0 + 1/2 \sum_{i=1}^{k} b_i (\bar{X}_{i1} + \bar{X}_{i2})$$

$$< b_0 + 1/2 \; [b_1 (\bar{X}_{11} + \bar{X}_{12}) + b_2 (\bar{X}_{21} + \bar{X}_{22})$$

$$+ - - - + b_k (\bar{X}_{k1} + \bar{X}_{k2})]$$

is classified into category two (for example don't own an automobile). In (4.74)and (4.75) $\bar{X}_{i1}$ and $\bar{X}_{i2}$ refer to means of the $i^{\underline{th}}$ predetermined variable for category one and category two, respectively. In other words the discriminating value is the estimated intercept in (4.73) plus one-half of the sum of the arithmetic means of the predetermined variables for the two categories weighted by the respective estimated regression coefficients from (4.73).

A coefficient of determination, R^2 and a coefficient of determination adjusted for degrees of freedom, $\bar{R}^2$, may be computed for a two category discriminate function but because of the non-normality of the disturbance term and the heterioscadastic nature of the variances hypotheses about goodness of fit and about individual regression coefficients cannot legitamately be tested.[2/]

[2/] Two references on discriminate analysis in which you might be interested are: 1) George W. Ladd, "Linear Probability Functions and Discriminate Functions", Econometrica, Vol. 34, No. 4, October 1966, pp. 873-885. and 2) George W. Ladd, Analysis of Ranking of Dairy Bargaining Cooperative Objectives, Iowa State University Research Bulletin, No. 550, January 1967, Ames, Iowa, pp. 887 - 893.

The least squares regression framework is applicable to discriminate analysis problems if the endogenous variable consists of only two classes, occurrence or nonoccurrence, of an attribute. If the endogenous variable consists of more than two classes ordinary least squares is not an appropriate estimation method and other multivariate techniques need to be employed.

Chapter 5 - Autocorrelation and Serial Correlation

Introduction

Autocorrelation and serial correlation are concerned with the lack of independence of successive distrubances. The successiveness of values may be based on a criterion such as ordering according to size of some exogenous variable or geographical proximity or in the case of serial correlation the natural sequencing over time. Autocorrelation and serial correlation refer to the same phenomena with the serial correlation referring to the interdependency over time.

Autocorrelation or serial correlation problems arise because the assumption $E(\varepsilon\varepsilon') = \sigma^2 I$ underlying the least squares estimation procedure is being violated. This assumption consists of two parts. The first part is that the σ^2 on the diagonal are all the same, hence, a common variance and second that $E(\varepsilon_i \varepsilon_r') = 0$ for $i \neq r$ or that off-diagonal elements are zero. In the case of autocorrelation or serial correlation $E(\varepsilon_i \varepsilon_r) \neq 0$ for $i \neq r$. If we specifically introduce time and focus on time series, then serial correlation exists when $E(\varepsilon_t, \varepsilon_{t+s}) \neq 0$ for all t and for some $s \neq 0$.

There are two main sources of serial correlation. First, the serial correlation may be due to a variable or set of variables omitted from the equation. Second, the serial correlation may be attributsble to a systematic measurement error in the endogenous model.

The omission of critical variables in an equation constitutes a model specification error which may or may not result in serial correlation of the disturbance terms. Just because a variable or set of variables are omitted from an equation doesn't mean that the disturbance will be serially correlated. Models generally are simplified representations

of a mechanism believed to have generated some set of observations and some variables are always omitted from the specification of the model. The influences of the omitted variables may work against each other in such a manner that there is no pattern in their combined behavior. In other words the effects of the individual components of the set of omitted variables cancel each other out and there would be no serial correlation due to omitted variables. It is only when the influence of certain of the omitted variables becomes over riding and exhibits a pronounced pattern of behavior that serial correlation problems arise. When this occurs, the disturbance term consists of two parts - a random element and a non-random component consisting of the combined effects of omitted variables.

Measurement error in the endogenous variable also may or may not lead to a serial correlation problem. If the measurement error is random and has an expected value of zero then no serial correlation would occur from the measurement error. It is only when the measurement error exhibits a systemmatic pattern and none of the exogenous variables account for the systemmatic pattern that the disturbances are serially correlated. Again the disturbance terms consists of two components - a random element and a systemmatic part. It is the systemmatic component which gives rise to the serial correlation.

Regardless of source, if something is to be done about the serially correlated disturbances either the basic forces responsible for the serial correlation need to be removed or adjustment made for the serial correlation either by introducing some scheme which will account for the serial correlation or which will adjust data and/or estimators for the serial correlation.

Simple Illustration of Serial Correlation

Let us consider serially correlated disturbances which follow a first order autoregressive scheme. In other words the disturbance is a linear function of its own lagged value as follows:

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + \delta_t \tag{5.1}$$

where ε_t is the disturbance in time t and δ_t is a random component in time t. Of course, more complex assumptions can be made about the autoregressive mechanism such as second, third, or higher orders, however, there existence in economic models may be questionable.

If in (5.1) $|\rho| > 1$ the autoregressive scheme is explosive and ε_t won't have a finite or constant variance. Only if $|\rho| < 1$ is the variance of ε_t a finite constant. In this latter case ε_t in some proportion of the distrubances in the preceding time period. Let us assume that the serial correlation coefficient ρ is $|\rho| < 1$ and that δ_t in (5.1) satisfies the assumptions

$$E(\delta_t) = 0 \tag{5.2}$$

and

$$E(\delta_t, \delta_{t+s}) = \sigma^2 \text{ for } s = 0 \qquad \text{for all } t$$
$$= 0 \text{ for } s \neq 0 \tag{5.3}$$

Hence, δ_t is the random component of the disturbance which satisfies the conditions which usually are assumed for least squares estimation. Now ε_t may be expanded in a series as follows:

$$(5.4) \qquad \varepsilon_t = \rho\,\varepsilon_{t-1} + \delta_t$$

$$= \rho\,(\rho\varepsilon_{t-2} + \delta_{t-1}) + \delta_t$$

$$= \rho^2\,(\rho\varepsilon_{t-3} + \delta_{t-2}) + \rho\delta_{t-1} + \delta_t$$

$$\vdots$$

$$= \rho^{\infty}\,\varepsilon_{t-\infty} + \delta_t + \rho\,\delta_{t-1} + \rho^2\,\delta_{t-2} + \cdots$$

$$= \rho^{\infty}\,\varepsilon_{t-\infty} + \sum_{s=0}^{\infty} \rho^s\,\delta_{t-s}$$

The first term of the last expression of (5.4) may be considered equal to zero without any difficulty since this amounts to changing the value of the constant term in the equation that has ε as its disturbance. Furthermore, any given ε, say $\varepsilon_{t-\infty}$, is anticipated to be small and has an expectation of zero in addition to the fact that $\rho^{\infty} \to 0$ since $|\rho| < 1$. Now

$$(5.5) \qquad E(\varepsilon_t) = 0 \text{ since } E(\delta_t) = 0 \text{ for all } t \text{ and furthermore,}$$

$$E(\varepsilon_t^2) = E(\delta_t^2) + \rho^2 E(\delta_{t-1}^2) + \rho^4 E(\delta_{t-2}^2) + - - -$$

since the δ's are serially independent and:

$$(5.6) \qquad E(\varepsilon_t^2) = (1 + \rho^2 + \rho^4 + \rho^6 + - - - +)\,\sigma_\delta^2$$

Based on the sum of an infinite geometric progression the variance of ε_t in (5.6) becomes

(5.7) $$\sigma_\varepsilon^2 = \frac{\sigma_\delta^2}{1 - \rho^2} \qquad \text{for all } t. \ \underline{1/}$$

Now

(5.8) $$\begin{aligned} E(\varepsilon_t, \varepsilon_{t-1}) &= E[(\delta_t + \rho\delta_{t-1} + \rho^2\delta_{t-2} + \text{---})(\delta_{t-1} + \rho\delta_{t-2} + \text{---})] \\ &= E[\delta_t(\delta_{t-1} + \rho\delta_{t-2} + \text{---} + (\delta_{t-1}\rho\delta_{t-2} + 0 \text{---})^2] \\ &= E[\delta_t(\delta_{t-1} + \rho\delta_{t-2} + \text{---})] + \rho E[(\delta_{t-1} + \rho\delta_{t-2} + \text{---})^2] \\ &= \rho\sigma_\varepsilon^2 \end{aligned}$$

Similarly:

(5.9) $$E(\varepsilon_t, \varepsilon_{t-2}) = \rho^2 \sigma_\varepsilon^2$$

and

(5.10) $$E(\varepsilon_t, \varepsilon_{t-3}) = \rho^3 \sigma_\varepsilon^2$$

1/Note that σ_ε^2 may be obtained directly from (5.1) as follows:

$$\begin{aligned} E(\varepsilon_t^2) &= E[\rho\varepsilon_{t-1} + \delta_t)^2] + E[\rho^2\varepsilon_{t-1}^2 + 2\rho\varepsilon_{t-1}\delta_t + \delta_t^2] \\ &= \rho^2 E(\varepsilon_{t-1}^2) + 2\rho E(\varepsilon_{t-1})\, E(\delta_t) + E(\delta_t^2) \\ &= \rho^2 \sigma_\varepsilon^2 + \sigma_\delta^2 \end{aligned}$$

Hence

$$\begin{aligned} \sigma_\varepsilon^2 &= \rho^2 \sigma_\varepsilon^2 + \sigma_\delta^2 \\ (1 - \rho^2)\sigma_\varepsilon^2 &= \sigma_\delta^2 \\ \sigma_\varepsilon^2 &= \frac{\sigma_\delta^2}{1 - \rho^2} \end{aligned}$$

In general

$$(5.11) \qquad E(\varepsilon_t, \varepsilon_{t-s}) = \rho^s \sigma_\varepsilon^2 \quad \text{for } s \neq 0$$

and

$$(5.12) \qquad \rho s = \frac{E(\varepsilon_t, \varepsilon_{t-s})}{\sigma_\varepsilon^2}$$

where $\frac{E(\varepsilon_t, \varepsilon_{t-s})}{\sigma^2}$ defines the sth order autocorrelation (serial) correlation coefficient of the ε series.

Effects of Serial Correlation

If ordinary least squares methods are applied when the disturbances are serially correlated there are three main consequences.

First, the estimators are unbiased, but the sampling variances of these estimators may be large compared with the sampling variances of estimators obtained by using different methods of estimation. The unbiased property follows directly from (5.5) since $E(b) = \beta + (X'X)^{-1} X'\varepsilon$. The ordinary least squares method is not taking the autoregressive nature of the disturbances into consideration and hence the least squares estimators are not minimal variance estimators. This suggests that estimation procedures that will really take into account the serial correlation and in some way adjust for it either by elimination or partial reduction should be more efficient estimators where more efficient means estimators with smaller sampling variances.

A second critical effect is that if the least squares formulas for sampling variances of the least squares estimators of regression coefficients are applied the sampling variances are likely to be seriously underestimated. There are two possible reasons why this may occur.

First, the least squares estimator of the sampling variance of regression coefficients ignores a set of terms which entail the serial correlation coefficient and the set of exogenous variables in the regression equation and which arise because the serial correlation coefficient is not equal to zero. Second, the estimated residual variance may be biased (understated) if the disturbances and hence the residuals are serially correlated and the serial correlation is ignored. This tends to be true if one or more of the exogenous variables and the disturbance term are serially correlated. Evolving directly from the problem of the sampling variances being underestimated by ordinary least squares methods, is that the conventional tests, t and F, used in testing various regression related hypotheses are no longer strictly valid.

A third effect is if predictions are made they will tend to be inefficient in the sense that they will have larger sampling variance then could be obtained if estimation and prediction were made by other methods which take the serial correlation into consideration. This is directly related to the first effect mentioned.

Identification of Serial Correlation Problems

In serial correlation problems, one is concerned with the interdependency among the disturbances; however, the disturbances are not observable so the identification of serial correlation has to be done through an examination of residuals which are observable. One of the simplest but less rigorous ways of analyzing the residuals is to plot the residuals and examine the curve for patterns which would suggest whether or not the residuals might be either positively or negatively serially correlated. Often this is a useful and expedient way of assessing whether serial correlation problems may exist. If the curve

resulting from the plotting of the residuals shows a trend from negative to positive or vice-versa or if it exhibits fairly smooth long waves with turning points being less common than continuations of either upward or downward movements, then very likely the residuals are positively serially correlated. Negatively serially correlated residuals will tend to exhibit persistent alternations between negative and positive values.

There are two statistical tests which are commonly used for testing hypotheses about serial correlation - von Neumann-Hart and Durbin-Watson tests. The von-Neumann-Hart test based on Hart's table is not strictly appropriate for testing hypotheses concerning serial correlation of disturbances in an equation because it was developed to apply to observable random variables. The distrubances are never observable so in order to make conclusion about the disturbances observable residuals which evolve from an estimated or calculated equation have to be used. The von Neumann-Hart test when applied to residuals of an equation tends to be biased toward the mean of the testing distribution and hence the hypothesis of no serial correlation is accepted too often when it should be rejected. That the Hart table is not appropriate was suggested by Monte Carlo sampling experiments conducted by Orcutt and Cochrane.[2/] They used artificial models with known serial correlation of shifts and known parameters to generate artificial time-series data. Sample sizes of $n = 20$ of this artificial data were used to estimate the parameters of the model. In a large number of the samples the hypothesis of no serial correlation was not rejected using the Hart table when it should have been rejected that it shatters the confidence one might have in the appropriateness of this test to residuals of an equation. The hypothesis of no serial correlation tended not to be rejected more often when in fact

---

[2/] See Carl Christ, Econometric Models and Methods, pages 524-525.

it should have been rejected when the number of variables in the equation were large.

The von-Neumann-Hart test is similar to the Durbin-Watson test which is an appropriate test for testing hypotheses about disturbances of an equation and it is presented mainly to show the relationship between the two tests. The von Neumann-Hart testing statistic is

$$(5.13)\quad \frac{\theta^2}{s^2} = \frac{\sum_{2}^{n} (X_t - X_{t-1})^2 /n-1}{\sum_{1}^{n} (X_t - \bar{X})^2 /n}$$

where:

X_t are random drawings from a normal distribution.

n is sample size, i.e., number of observations.

The $\frac{\theta^2}{s^2}$ distribution was developed by von Neumann and tabulated by Hart.

The rejection and non-rejection regions of the von-Neumann-Hart test are presented in Figure 5.1.

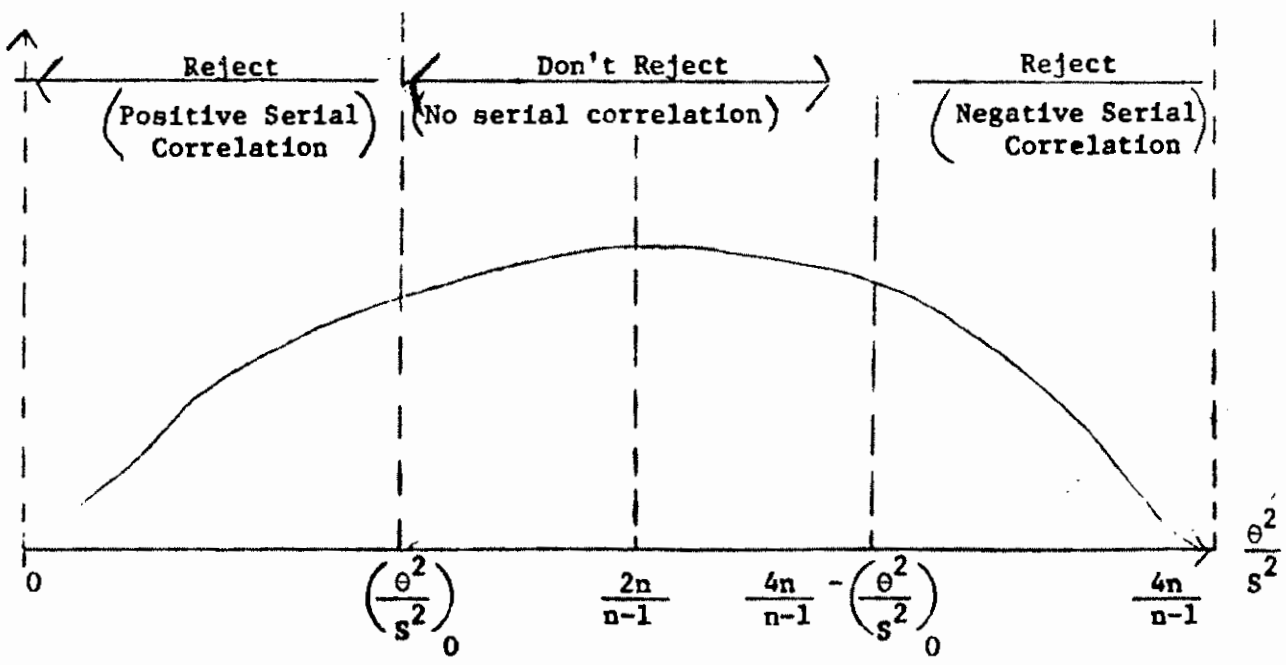


Figure 5.1. Diagramatic representation of von Neumann-Hart test for serial correlation.

The distribution of $\frac{\theta^2}{s^2}$ has a mean value of almost 2 and computed values of the testing statistic near 2 indicate little or no serial correlation. If the random variable exhibits smooth upward or downward trend or a smooth wavy pattern then $\frac{\theta^2}{s^2}$ will take on small values which indicate positive serial correlation. If a high value of the random variable is usually followed by a low value and a low value followed by a high value etc. then $\frac{\theta^2}{s^2}$ will take on large values indicating a negative serial correlation among the random variables. To apply the test, one needs the table of values of $\frac{\theta^2}{s^2}$ as presented in Table B-4 of Christ (1966).

The Durbin-Watson test which is similar to the von Neumann-Hart test is appropriate for testing hypotheses about serial correlation of disturbances of a regression equation. The test applies when all of the independent or predetermined variables are strictly exogenous. In general it is not appropriate to equations which belong to a simultaneous system or to equations that contain lagged values of the endogenous variable as an "exogenous" variable.

The Durbin-Watson statistic, d, is almost the same as $\frac{\theta^2}{s^2}$ but differing by the factor (n/n-1).

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \tag{5.14}$$

Remember that $\sum_{t=1}^{n} e_t = n\bar{e} = 0$ so that $\sum_{t=1}^{n} (e_t - \bar{e})^2 = \sum_{t=1}^{n} e_t^2$

The rejection and non-rejection regions of the Durbin-Watson statistic are presented in diagramatic form in Figure 5.2.

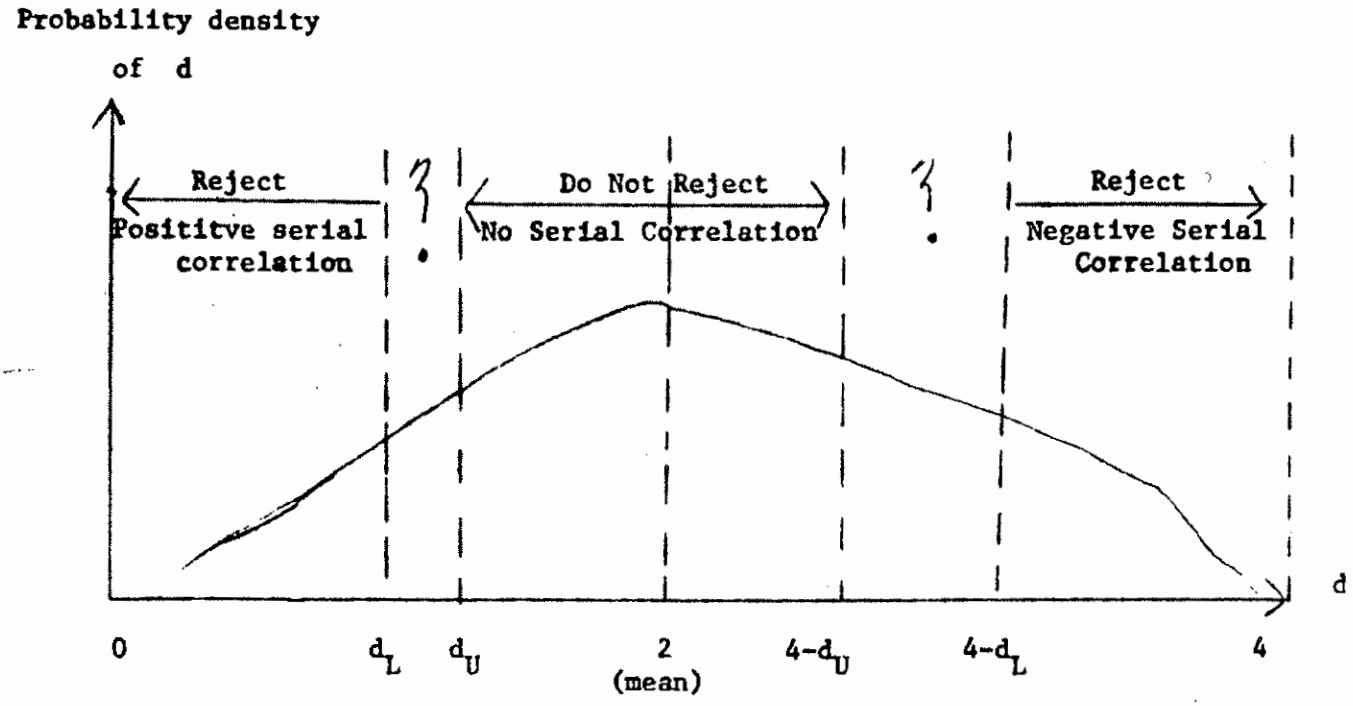


Figure 5.2. Diagramatic representation of Durbin-Watson test for serial correlation.

In testing for positive serial correlation the following limits and conclusions apply:

1. If d is below d_L reject the null hypothesis of no serial correlation and conclude that positive serial correlation exists.

2. If d is above d_U do not reject the null hypothesis of no serial correlation and conclude that there are no serial correlation.

3. If d is between d_L and d_U the test is inconclusive.

In testing for negative serial correlation the following limits and conclusions apply.

1. If d is above $4 - d_L$ reject the null hypothesis of no serial correlation and conclude that negative serial correlation exists.

2. If d is below $4 - d_U$ do not reject the null hypothesis of no serial correlation and conclude that there is no serial correlation.

3. If d is between $4 - d_L$ and $4 - d_U$ the test is inconclusive.

The Durbin-Watson test may be applied to cross-section residuals arranged in any nontemporal order to test hypotheses about autocorrelation. In equations containing the lagged endogenous variable as a predetermined variable the Durbin-Watson test is biased against identifying either positive or negative serial correlation. The Durbin-Watson test appears to be an extremely good approximation for a power function in which the exogenous variables are various powers of a given exogenous variable, i.e., $Y = f(X, X^2, X^3, X^4, \text{etc.})$. In the case of simultaneous systems of equations, the Durbin-Watson test applies to the residuals of an equation that is just identified and it is a good approximate test for the residuals of an over-identified equation where all predetermined variables are assumed to be strictly exogenous, i.e., no lagged dependent variables as predetermined variables.

3/ For more details on the topics presented in this paragraph see Carl Christ, Econometric Models and Methods, pages 527-531.

Some Difficulties in Testing for Serial Correlation

One of the major difficulties in testing whether or not the disturbances of a regression model are uncorrelated is that the disturbances are not observable. The common procedure in testing hypotheses about the disturbances is to use the least squares residuals. The residuals are:

$$(5.15) \qquad e = (I - X(X'X)^{-1} X')\varepsilon$$

where X is the matrix of observations of the predetermined variables

ε is the vector of true disturbances.

$(I - X(X'X)^{-1} X')$ is idempotent, symmetric, and positive semi-definite and its rank is $n - k - 1$ where n is sample size and $(k + 1)$ are the degrees of freedom "lost" (k predetermined variables plus the vector of ones for the intercept). $(I - X(X'X)^{-1} X')$ is the matrix that transforms the true disturbances, ε, linearly into the estimated residuals.

Note that the residuals depend not only on ε but also on the values of the predetermined variables, X through $(I - X(X'X)^{-1} X')$. This implies that even if

$$(5.16) \qquad E(e) = 0 \text{ and } E(\varepsilon\varepsilon') = \sigma^2 I$$

are true the residuals , e, are in general not uncorrelated nor do they have constant variance. The variance-covariance matrix of the residuals is:

$$(5.17) \qquad E(ee') = E\left[(I - X(X'X)^{-1} X')\ (I - X(X'X)^{-1} X')'\right]$$

$$= \sigma^2 \left[(I - X(X'X)^{-1} X')\ (I - X(X'X)^{-1} X')'\right]$$

$$= \sigma^2\ (I - X(X'X)^{-1} X')$$

Remember $(I-X(X'X)^{-1} X')$ is symmetric and equal to its own square. This complicates matters in terms of testing. The Durbin-Watson procedure for testing against serial correlation is formulated in terms of upper and lower bounds to significance limits. Although it is possible to simplify the testing procedure under certain conditions in most cases the testing procedure is not as simple as one might get the impression from the Durbin-Watson testing procedure. Also, the testing procedure for alternative hypotheses such as heteroscadasticity and non-normality of disturbance would be simplified considerably if the variance-covariance matrix of the residuals were of the form $\sigma^2 I$ rather than $\sigma^2(I - X(X'X)^{-1} X')$.[4/]

There are some aspects of the Durbin-Watson test which need to be pointed out. Criticisms have been directed to these aspects and alternative tests have been suggested.[5/] One of these aspects is the inconclusive region of the Durbin-Watson test. This is a region of ignorance. This region is large when the number of observations is small or moderate (around 20) and the number of explanatory variables not so small.[6/] This is not particularly helpful to the applied researcher who finds a lot of the tests he performs in the inconclusive region and thus he has no guidance in terms of whether there is or is not serial correlation. Although

---

[4/] See the article by H. Theil, "The Analysis of Distrubances in Regression Analysis", JASA, Vol. 60, No. 312, December 1965, pp. 1067-1079. The above discussion has drawn on the first two pages of this article. In this article Theil develops his BLUS (Best Linear Unbiased Scaler Covariance Matrix) estimators.

[5/] See H. Theil and A. L. Nagar, "Testing the Independence of Regression Disturbances", JASA, Vol. 56, No. 296, December 1961, pp. 793-806. Richard C. Henshaw, Jr., "Testing Single-Equation Least Squares Regression Model for Autocorrelated Disturbances", Econometrica, Koteswara Rao Kadiyala, "Testing for the Independence of Regression Disturbances", Econometrica, Vol. 38, No. 1, January 1970, pp. 97-117.

[6/] Theil and Nagar, Ibid., pp. 793.

many researchers interpret the "no inference possible" as "no need to reject the null hypothesis of independence among disturbances" and proceed as if there was no serial correlation. Generally it is pointed out that certain tests were inconclusive but then the equations are grouped with those for which the test indicated no serial correlation and no distinction is made in interpretation and inferences from there on. There is a danger in this that it will lead to bias in the sense that too many cases of positive serial correlation are overlooked

One aspect of the Durbin-Watson test which has come under fire is the theoretical underpinnings of the test. The inconclusive regions of the test are not necessarily independent of this. The difficulties with this underlying theory may be set forth from the Durbin-Watson statistic presented in (5.14) which is:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{e'A\,e}{e'e} \tag{5.18}$$

where:

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdots & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & \cdots & & & & 2 & -1 \\ & & & & & -1 & 1 \end{bmatrix} \tag{5.19}$$

The distribution of d depends on the particular set of regressors in the data under analysis and varies from sample to sample.[7/] In other words the distribution of d depends on the particular X matrix being considered. This means that there is a different distribution of d for each set of predetermined variables. But the rejection values, d_L and d_U, of the Durbin-Watson test are general bounds in the sense that the particular X matrix is not taken into consideration although sample size and the number of predetermined variables are. So the Durbin-Watson test is an approximate test which does not deal with the exact distribution of d for the particular X matrix. Under certain restrictive conditions the bounds established by Durbin-Watson are known to hold. Under the null hypothesis of independence and normal distribution of the disturbances if the columns of X as spanned by the eigenvectors of A corresponding to the k smallest characteristic roots of A the upper bound of d is appropriate. Under the normality and independence assumptions if the columns of X are spanned by the eigenvectors of A corresponding to the k largest characteristic roots of A the lower bound of d is appropriate. However, a question arises concerning the justification of the assumption that the columns of X are linearly dependent on some k eigenvectors of A. It is not clear why this should be and furthermore it is not clear how this may vary for different X matrices. Even more serious questions arise for

---

[7/] See J. Durbin, "An Alternative to the Bounds Test for Testing for Serial Correlation in Least-Squares Regression", Econometrica, Vol. 38, No. 3, May 1970, pp. 422-429. In this particular article Durbin shows how to transform residuals from regression on an arbitrary set of k regressors to a set of values having the joint distribution as the residuals from regression on a different set of L of k regressors. He denotes that $d' = \frac{\Sigma(e_t - e_{t-1})^2}{e_t^2}$ calculated from these values. He shows that for a suitable choice of L the distribution of d' is the same as d_U, the significance values tabulated for the Durbin-Watson statistic. Suitable choice still gives rise to some difficulties.

the conditions of the alterntive hypothesis that the disturbances are not independent and/or they are not normally distributed. It is not at all clear what the implications of applying the Durbin-Watson test under these conditions are and it is not apparent that is an appropriate test. In addition the question of what happens to the d distribution for different X matrices continues to persist.

Theil and Nagar[8/] pointed out that the regressor used in much time-series regression are changing slowly in the sense that their differences of orders one to four are small in relation to the ranges of the variables themselves. Using this assumption they developed an alterntive test based on the moments of d where d is the Durbin-Watson statistic defined in (5.18). The Theil-Nagar test is based on the first four moments of the d distribution which is assumed to be approximated by a Beta distribution. Because of the slowly changing nature of the regressors, certain components of the first four moments are dropped. This makes the computation procedure less complicated but their test is only an approximation

Henshaw[9/] modified the Theil-Nagar test by improving the accuracy of fitting a Beta approximation based on the values of the first four moments of d. The method requires a substantial amount of computation however, with modern day computers this is not such a burdensome task. The Henshaw test is based on the fact that in least squares regression the attainable limits of d are determined by the roots of the matrix

$$K = A - AX\,(X'X)^{-1}\,X' \tag{5.20}$$

where A is defined in (5.19). (5.20) is positive semi-definite with (k+1) zero roots. The range of d consists of the smallest and largest of the (n - k - 1) positive roots of K. The numerical values of the first four

---

8/ Theil and Nagar, op. cit.

9/ Henshaw, op. cit., pp. 648-651

moments of the d distribution are calculated from the data as follows:

$$(5.21) \qquad E(d) = \frac{1}{[n-k-1]} \cdot \text{tr } K$$

$$(5.22) \qquad \mu_2 = \text{var } (d) = 2 \left\{ \frac{[n-k-1] \text{ tr } K^2 - (\text{tr } K)^2}{[n-k-1]^2 [n-(k+1)+2]} \right\}$$

$$(5.23) \qquad \sqrt{\beta_1} = \frac{\mu_3}{\mu_3^{3/2}} = 8 \left\{ \frac{[n-k-1]^2 \text{ tr } K^3 - 3[n-k-1] \text{tr } K \text{tr} K^2 + 2(\text{tr} K)^3}{[n-k-1]^3 [n-(k+1)+2][n-(k+1)+4] \mu_2^{3/2}} \right\}$$

$$(5.24) \qquad \beta_2 = \frac{\mu_4}{\mu_2} = 12 \{[n-k-1]^3 (4 \text{ tr } K^4 + (\text{tr} K^2)^2] - 2 [n-k-1]^2 \text{x} [8 \text{tr} K \text{tr} K^3 + \text{tr} K^2 (\text{tr} K)^2] + [n-k-1][24 \text{tr} K^2 (\text{tr} K)^4] - 12 (\text{tr} K)^4\} + \{[n-k-1]^4 [n-(k+1)+2] \{n-(k+1)+4] [n-(k+1)+6] \mu_2^2 \}$$

k is the number of predetermined variables excluding the vector of 1's for the intercept, i.e., (k+1) corresponds to Henshaw's k and Theil and Nagar's A. The results (5.21) - (5.24) are then used to fit the Beta function. They are exact and may be computed for any k.

A Beta variable in the range (0,1) has the following probability density function:

$$(5.25) \qquad f(w; p,q) = \frac{1}{\beta(p,q)} \cdot X^{p-1} (1 - X)^{q-1}$$

where

$$\frac{1}{\beta(p,q)} = \frac{(p+q-1)!}{(p-1)! (q-1)!}$$

The first four moments of the Beta function are:

$$(5.26) \qquad E(w) = \frac{p}{p+q}$$

$$(5.27) \quad \text{var } (w) = \frac{pq}{(p+q)^2 \ (p+q+1)}$$

$$(5.28) \quad \sqrt{\beta_1} = \frac{2(q-p) \ (p+q+1)^{1/2}}{(p+q+2) \ (pq)^{1/2}}$$

$$(5.29) \quad \beta_2 - 3 = \frac{6[(p+q+1) \ (q-p)^2 - (p+q+2) \ (pq)]}{pq \ (p+q+2) \ (p+q+3)}$$

β_1 and (β_2-3) are independent of origin and scale. (5.28) and (5.29) may be solved for p and q in terms of β_1 and β_2. (5.23) and (5.24) may be used to estimate β_1 and β_2 from the data. In order to solve for p and q in terms of β_1 and β_2 let:

$$(5.30) \quad M = \frac{4 \beta_2 - 3 \beta_1}{3 \beta_1 - 2 \beta_2 + 6}$$

and

$$(5.31) \quad N = \frac{\sqrt{\beta_1} \ (M^2 - 1)}{[16M + \beta_1 \ (M+1)^2]^{1/2}}$$

The solutions to p and q are:

$$(5.32) \quad p = 1/2 \ (M - N - 1)$$

and

$$(5.33) \quad q = 1/2 \ (M + N - 1)$$

Now the d statistic may be written (λ_s, λ_g) where λ_s and λ_g are the smallest and greatest non-zero characteristic roots of K, respectively. d can be transformed into a beta variate w in the range (0,1) in the following manner:

$$(5.34) \quad w = \frac{d - \lambda_s}{\lambda_g - \lambda_s}$$

with

$$(5.35 \qquad E(w) = \frac{E(d) - \lambda_s}{\lambda_g - \lambda_s}$$

and

$$(5.36) \qquad \text{var}(w) = \frac{\text{var}(d)}{(\lambda_g - \lambda_s)^2}$$

The solutions for λ_s and λ_g are:

$$(5.37) \qquad \lambda_s = E(d) - E(w) \left[\frac{\text{var}(d)}{\text{var}(w)}\right]^{1/2}$$

and

$$(5.38) \qquad \lambda_g = \left[\frac{\text{var}(d)}{\text{var}(w)}\right]^{1/2} + \lambda_s$$

which may be calculated from the data by substituting E(d), var (d), E(w), and var (w) from (5.21), (5.22), (5.23), and (5.24), respectively.

To estimate the significance points of d the relationship between w and d implied by (5.34) may be used: $(5.39) \qquad d = \lambda_s + (\lambda_g - \lambda_s)\, w$

First, p and q as estimated by (5.32) and (5.33) should be replaced by tabled values nearest them in the Tables of the Incomplete Beta-Function.[10/] so that values of w corresponding to the desired levels of significance can be determined. Interpolation should be used for more accuarte results. Substituting such values of w into (5.39) yields the required significance points of d, i.e., $d_{.05}$ or $d_{.01}$ etc.

[10/] Pearson, K., Tables of the Incomplete Beta-Function, Cambridge, Biometrika Office, 1934.

There are other tests for serial correlation all of which entail some rather burdensome calculations. Also, some procedures for circumventing the problem have been suggested. For further information see the reference indicated in the footnotes of this section.[11/]

11/ J. Durbin, op. cit.
Hannan, E., Jr. and R. D. Terrel, "Testing for Serial Correlation After Least Squares Regression", Econometrica, Vol. 36, 1968, pp. 133-155.
Kadiyala, Koteswara R., "Testing for Independence of Regression Disturbances", Econometrica, Vol. 38, 1970, pp. 97-117.
Kadiyala, Koteswara R., "A Transformation Used to Circumvent the Problem of Auto Correlation", Econometrica, Vol. 36, 1968, pp. 93-96.

Alternative Approaches for Handling Serial Correlation Problems

There are several ways of tackling the serial correlation problem. Some involve adjustments utilizing outside information, some utilize information computed from the sample data, others involve modifications of the basic model. Each of several approaches will be discussed briefly:

Reformulate Model - Include Excluded Variables

One of the causes of serially correlated disturbances is the omission of a variable or set of variables which are crucial in accounting for the variation in the endogenous. If the effects of the omitted variables is overriding and exhibit a systemmatic pattern the disturbance consists of two components, the systemmatic one and the random element. Including one or more of the key omitted variables could reduce or eliminate the serial correlation among the observed residuals.

The major difficulty with this approach is identifying one or more such exogenous variables that have a theoretical justification for being included in the model. If a thorough and rigorous job was performed in formulating the model initially, it is not necessarily easy to identify specific variables which might be candidates to be included in the model to alleviate the serial correlation problem. At the same time, frequently in formulating a model, the desirability of a simplified model leads one to exclude variables which had a basis for being in the model but whose inclusion would make the model more complex and more difficult to estimate and analyze. These may be prime candidate to which to turn. Also, in the initial formulation frequently one has to choose between alternative functional forms, all of which may be theoretically justifiable, alternative measures of a given variable and alternative sets of variables. Some of all of these could be the source of the serial correlation problem and thus be areas where solutions might be found.

A high degree of serial correlation among the residuals is a signal to carefully scrutinize the theory used in formulating the model. There is such a thing as using a good theory in the wrong way at the wrong time to formulate a model with serial correlation and other types of difficulties. After one has the information that serial correlation exists often one has no difficulty in enumerating reasons why knowledge that serial correlation exists is not astounding in that particular situation. On rare occasions, one may even use alternative theories to formulate a different model for the same problem situation and eliminate the serial correlation problem.

Iterative Estimation of ρ.

The first step in this procedure is to apply ordinary least squares to the data as if the disturbances were not serially correlated. Then compute the residuals and denote them as e_t^0. Next using least squares techniques estimate the autoregression coefficient of the relationship.

(5.40) $$e_t^0 = \hat{\rho}_1 e_{t-1}^0 + \hat{\delta}_t^0$$

where:

(5.41) $$\hat{\rho}_1 = \frac{\frac{1}{n-1} \sum_{t-2}^{n} e_t^0 e_{t-1}^0}{\frac{1}{n-k-1} \sum_{t=1}^{n} (e_t^0)^2} = \frac{\sum_{t=2}^{n} e_t^0 e_{t-1}^0}{(n-1)\, s^2}$$

Next transform each variable, endogenous and all exogenous variables, into a new set of variables consisting of the raw data minus $\hat{\rho}_1$ times the lagged value of each variable.

(5.42) $Y_t - \hat{\rho}_1 Y_{t-1}$ where Y_t and Y_{t-1} are (n-1) x 1 vectors

and

(5.43) $X_t = \hat{\rho}_1 X_{t-1}$ where X_t and X_{t-1} are (n-1) x (k+1) matrices. Now re-estimate the regression equation using the transformed data based on $\hat{\rho}_1$. Then compute the residuals and denote them as e_t'. A test regarding serial correlation in these residuals may be performed and if they are not

significantly serially correlated the iterative procedure is stopped. If the residuals, e'_t, are still highly correlated we can estimate the autoregression coefficient of the relationship:

$$(5.44) \qquad e'_t = \hat{\rho}_2 \, e_{t-1} + \hat{\delta}'_t$$

where

$$(5.45) \qquad \hat{\rho}_2 = \frac{\frac{1}{n-1} \sum_{t=2}^{n} e'_t \, e'_{t-1}}{\frac{1}{n-k-1} \sum_{t=1}^{n} (e'_t)^2} = \frac{\sum_{t=2}^{n} e'_t \, e'_{t-1}}{(n-1)\,(s')^2}$$

Now transform the original data once again using $\hat{\rho}_2$ instead of $\hat{\rho}_1$. The transformed data are:

$$(5.46) \qquad Y_t = \hat{\rho}_2 \, Y_{t-1} \text{ where } Y_t \text{ and } Y_{t-1} \text{ are } (n-1) \times 1 \text{ vectors}$$

and

$$(5.47) \qquad X_t = \hat{\rho}_2 \, X_{t-1} \text{ where } X_t \text{ and } X_{t-1} \text{ are } (n-1) \times (k+1) \text{ matrices.}$$

Now re-estimate the regression equation using the transformed data based on $\hat{\rho}_2$. Then compute residuals and denote them as e''_t. Next test hypotheses about serial correlation in the residuals, e''_t, and if they are not significantly serially correlated the iterative estimation and testing procedure is stopped. If the residuals are still serially correlated, repeat the data transformation and estimation based on the newly transformed data. Follow this by computing the residuals of this most recently estimated relationship and test hypotheses about serial correlation for those residuals. Continue repeating this process until a set of residuals which are not serially correlated are obtained.

Assume a Value of ρ

Instead of trying to estimate the autoregression coefficient alternatively one may assume a value and transform the data similarly to the transformations in (5.46) and (5.47). Then one can estimate the regression

equation, compute residuals, and test hypotheses about serial correlation of the residuals. This is particularly a suitable approach if one has some prior information about the magnitude of the autoregression coefficient.

When one is not so fortunate as to have a prior estimate of ρ, then choosing a value of ρ to assume may be a rather difficult task. There is usually no reason to assume a value outside the range of -1 to +1 since values of ρ outside this range would be coefficients of explosive autoregressive schemes and these are not very interesting to economists. Usually we don't observe relationships shifting back and forth from one extreme to the other so that negative values of ρ are not very plausible. Hence, the most plausible range is from 0 to +1. $\rho = 0$ indicates no serial correlation and the shifts exert their influence in time period t and then are gone. On the other extreme, $\rho = 1$, indicates that the influence of shifts in time period t are permanent. The shifts in each successive time period are measured from the position established in the immediately preceding time period.

The case where it is assumed that $\rho = 1$ entails a special transformation on the data. The transformation is:

$$(5.48) \qquad Y_t - Y_{t-1} = \Delta Y_t \text{ and } X_t - X_{t-1} = \Delta X_t$$

These are first differences and a regression equation estimated on the basis of these data is the first difference equation. One thing to consider in using first differences is whether the phenomena being investigated is conformable to the change conditions implied by a first difference equation. Also, under certain condition first differences may introduce serial correlation where otherwise there would be no serial correlation.

Incorporate Lagged Variables

Many economic variables are serially correlated and incorporating both the current and lagged values of an exogenous variable or set of

exogenous variables in the regression equation creates a mechanical way of eliminating or reducing the serial correlation in the residuals. Hence, this mechanical device accounts for the serial correlation and it is not forced upon the disturbance term. It should be recognized that this is a mechanical procedure and should not be used if the correct or more appropriate model is known.

There are a couple of dangers about using both current and lagged values of exogenous variables in regression equations which need to be guarded against. First, in some types of relationships certain lagged variables may not be meaningful and it is questionable whether such should be used just to alleviate a serial correlation problem. Second, and probably more important, current and lagged values of a given variable usually tend to be highly correlated and including both as predetermined variables in an equation may lead to a severe multicollinearity problem. Hence, one may eliminate one class of problems only to create another class of problems. Of course, relevant issues with respect to this are the severity of the multicollinearity that was created and its consequences relative to the consequences of the serial correlation problem.

Use Knowledge of Variance-Covarianee Matrix

One of the more ideal ways and at the same time more impractical ways of handling the serial correlation problem is to incorporate knowledge about the variance-covariance matrix of the disturbance terms, ϕ, right into the estimation process. If the disturbance term follows a first-order autoregressive scheme then:

$$(5.49) \qquad \phi = \frac{\sigma_\delta^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 \rho & \cdots & \rho^{n-3} \\ & \cdot & & & \cdot \\ & \cdot & & & \cdot \\ & \cdot & & & \cdot \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}$$

The generalized least squares procedure needs to be used and the GLS estimators are

$$(5.50) \qquad \tilde{\beta} = (X' \phi^{-1} X)^{-1} X' \phi^{-1} Y$$

with variances

$$(5.51) \qquad \text{var } (\tilde{\beta}) = (X' \phi^{-1} X)^{-1}$$

The obvious difficulty with this procedure is that we usually don't know ϕ. If an estimate of ϕ, say ϕ_1, is available it can be utilized in estimation. The resulting estimators will be unbiased but there will be an inaccuracy in computing the variance-covariance matrix of the estimators. The better the estimate that ϕ_1 is of ϕ the less the inaccuracy. Also, a slightly inaccurate estimate of ϕ may be better than neglecting the serial correlation problem by applying ordinary least squares methods.

One estimate of ϕ which comes to mind immediately is to use the ordinary least squares residuals and then to use generalized least squares to re-estimate the coefficients. This would mean letting:

$$(5.52) \qquad b = (X'X)^{-1} X'Y$$

and

$$(5.53) \qquad e = Y - Xb$$

and

(5.54) $\phi_2 = ee'$

where

(5.55)
$$\phi_2 = \begin{bmatrix} e_1^2 & e_1e_2 & \cdots & e_1e_n \\ e_2e_1 & e_2^2 & \cdots & e_2^n \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ e_ne_1 & e_ne_2 & \cdots & e_n^2 \end{bmatrix}$$

However, this is not a workable method because ϕ_2 is singular and ϕ_2^{-1} does not exist. Note that

(5.56)
$$|\phi| = \begin{vmatrix} e_1^2 & e_1e_2 & \cdots & e_1e_n \\ e_2e_1 & e_2^2 & \cdots & e_2e_n \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ e_ne_1 & e_ne_2 & \cdots & e_n^2 \end{vmatrix} = e_1e_2 \cdots e_n \begin{vmatrix} e_1e_1 & \cdots & e_1 \\ e_2e_2 & \cdots & e_2 \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ e_ne_n & \cdots & e_n \end{vmatrix}$$

$= 0$ because the rows of the latter expression in (5.56) are linear combinations of each other i.e., $e_1 = a\ e_2$ etc. and columns are identical.

If one is dealing with a large sample there is always the possibility of splitting the sample into parts and using one part to obtain an estimate of ϕ and use the other part or parts to estimate the regression coefficients. However, the usual sample size with which economists are confronted is small and hence the estimate of ϕ will have to come from elsewhere than the sample data being analyzed if this method is to be used.

Simultaneous Estimation of All Parameters Including the Autoregression Coefficient

This method entails the use of maximum-likelihood estimation procedures. It involves making a transformation on the autoregressive residuals to transform them into a random variable with zero mean and constant variance. This transformation is incorporated into the estimation procedure. The procedure requires that the transformed residuals are assumed to be normally distributed. The estimating equations turn out to be nonlinear in unknowns and solving these is not only cumbersome but tedious and hard work.

Chapter 6 - Generalized Least Squares

Introduction

In the previous chapter it was indicated that generalized least squares was an appropriate method of estimation for incorporating information about the variance-covariance matrix of disturbances into the estimation of parameters of a linear multiple regression equation when serial correlation is present in the disturbances. In this section an attempt will be made to briefly develop some of the fundamentals of generalized least squares (GLS) estimators.

As one might guess from the name, generalized least squares, this method of estimation utilizes the criterion of a minimum sum of squared residuals in obtaining estimators of parameter of a linear multiple variable model. Furthermore, one would suspect that under certain conditions, the ordinary least squares procedure is probably a special case of generalized least squares. And this is true. The generalization of least squares is attributable to Aitken[1/] in that he established the generalization of the basic Gauss-Markov theorem underlying least squares. Some times the generalized least squares estimators are referred to as Aitken's estimators.

As a point of departure, let us consider a linear multiple variable model of the form

$$Y = X\beta + \varepsilon \tag{6.1}$$

where:

Y is $n \times 1$ vector of the endogenous variable.

X is $n \times (k+1)$ matrix of exogenous variables.

β is a $(k+1) \times 1$ vector of unknown parameters.

ε is a $n \times 1$ vector of disturbances which may or may not be random.

[1/] A. C. Aitken, "On Least Squares and Linear Combinations of Observations," Proceedings Royal Society, Edinburgh, Vol. 55, pp. 42-48, 1934.

Let us assume that

(6.2) $E(\varepsilon) = 0$

(6.3) $E(\varepsilon\varepsilon') = \phi$ where ϕ is a positive definite matrix of variances and covariances of the disturbance.

(6.4) X is n x (k+1) matrix of fixed numbers or variables which do not change values in repeated sampling.

(6.5) The rank of X is $(k+1) \leq n$.

Note that one of the major differences in assumptions underlying generalized least squares and ordinary least squares is (6.3). In ordinary least squares we assumed a scalar matrix, $\sigma^2 I$, where as (6.3) specifies a non-scalar matrix. This means that the variances on the diagonal of ϕ may not be the same and the covariances as off-diagonal elements of ϕ may be different from zero. Only infinite variances and singular ϕ are excluded. There are several reasons why we might have a variance-covariance matrix such as ϕ, one of which we have already discussed and that is the case of serial correlations of disturbances. One desirable aspect of the generalization of least squares is that estimators with desirable properties may be obtained when the variance-covariance matrix is not scalar but one as indicated in (6.3).

Now let us further explore the characteristics of ϕ when it is positive definite which we are assuming it is since it is a variance-covariance matrix. We may express the variance-covariance matrix as a product of the disturbance variance and another matrix such as

(6.6) $\phi = \sigma_\varepsilon^2 A$

where σ_ε^2 is positive and A is a positive definite matrix.

Note that for a positive autoregression coefficient (the case of positive serial correlation) that the matrix of ρ's is this type of matrix. If $E(\varepsilon\varepsilon') \neq \sigma_\varepsilon^2 I$ but equal to some ϕ as in (6.3) and X as expressed in (6.1) has full column rank, ordinary least squares methods yield estimators which are unbiased but they

are not best in the sense of being minimum variance estimators. For the generalized linear regression model, the best linear unbiased estimators (BLUE) of β is the generalized least squares vector

$$(6.7) \qquad \tilde{b} = (X'\phi^{-1}X)^{-1} X'\phi^{-1}Y = (X' A^{-1} X)^{-1} X' A^{-1}Y$$

with a variance-covariance matrix of:

$$(6.8) \qquad V(\tilde{b}) = (X'\phi^{-1}X)^{-1}$$

It is clear that when $\phi = \sigma_\varepsilon^2 I$ the Aitken's estimators given in (6.7) and (6.8) are:

$$(6.9) \qquad \tilde{b} = (X'\phi^{-1}X)^{-1} X'\phi^{-1}Y = \sigma_\varepsilon^2 (X'X)^{-1} \ 1/\sigma_\varepsilon^2 X'Y = (X'X)^{-1} X'Y$$

and

$$(6.10) \qquad V(\tilde{b}) = (X' \ 1/\sigma_\varepsilon^2 IX)^{-1} = \sigma_\varepsilon^2 (X'X)^{-1}$$

Now let us focus on how we may go about obtaining the Aitken's type of estimators.

One way of obtaining the best estimators under the conditions specified in (6.1)-(6.5) is to transform the variables, endogenous and all exogenous, in such a way that the variance-covariance matrix after the transformation is $\sigma_\varepsilon^2 I$ instead of $\phi = \sigma_\varepsilon^2 A$. To proceed with this we need to take advantage of some fundamental properties of symmetric positive definite matrices. If A is symmetric and positive definite then A^{-1} is also symmetric and positive definite and there exists a non-singular n x n matrix P such that $P'P = A^{-1}$.[2/] Now we have:

$$(6.11) \qquad \phi = \sigma_\varepsilon^2 A = \sigma_\varepsilon^2 (P'P)^{-1}$$

Remultiplying both sides of (6.11) by P we have

$$(6.12) \qquad PY = P X \beta + P \varepsilon$$

where

PY is the new vector of observations of the endogenous variable.

PX is the new matrix of the observations of the exogenous variables.

Pε is the new vector of disturbances.

[2/] For the proof of this see Arthur S. Goldberger, Econometric Theory, 1964, pages 35 and 36 particularly (7.6) and (7.11) in the section concerned with positive definite matrices.

We are choosing P such that it is independent of ε and hence

(6.13) $\quad E(P\varepsilon) = E(P)\ E(\varepsilon) = E(P)\ \ (0) = 0.$

Now

(6.14) $\quad E[(P\varepsilon)(P\varepsilon)'] = E(P\varepsilon\varepsilon'P') = P\ \phi\ P' = P\ \sigma_\varepsilon^2\ A\ P' = \sigma_\varepsilon^2\ P(PP')^{-1}\ P' = \sigma_\varepsilon^2\ \ I$

where $\sigma_\varepsilon^2\ P(PP')^{-1}\ P'$ is based on $P'P = A^{-1}$.

Applying the least squares criterion to (6.12) instead of (6.1) yields the generalized least squares estimators given in (6.7) and (6.8) namely

(6.15) $\quad \tilde{b} = (X'A^{-1}X)^{-1}\ X'\ A^{-1}\ Y = (X'\ P'P\ X)^{-1}\ X'\ P'PY$

and

(6.16) $\quad V(\tilde{b}) = \sigma_\varepsilon^2\ (X'A^{-1}\ X)^{-1} = \sigma_\varepsilon^2\ (X'\ P'P\ X)^{-1}$

In the ordinary least squares framework of estimation, the estimator of the residual variance is obtained by dividing the squared length of the least squares residual vector by the degrees of freedom associated with the sum of squared residuals, i.e., $s^2 = \frac{e'e}{n-k-1} = \frac{(Y - Xb)'\ (Y-Xb)}{n\ -\ k\ -\ 1}$. In the generalized least squares framework the quadratic form $(Y - X\tilde{b})'\ A^{-1}\ (Y - X\tilde{b})$ is divided by the degrees of freedom associated with these sum of squared residuals to obtain an estimate of the residual variance. The residual to be minimized is

(6.17) $\quad (P\tilde{e}) = P(Y - \tilde{Y}) = P(Y - X\tilde{b})$

where $\tilde{e} = (Y - \tilde{Y})$ and $\tilde{Y} = X\tilde{b}$. The estimator of the residual variance is:

(6.18) $\quad \tilde{\sigma}_\varepsilon^2 = \frac{(P\tilde{e})'(P\tilde{e})}{n-k-1} = \frac{\tilde{e}'\ P'\ P\tilde{e}}{n-k-1} = \frac{(Y - X\tilde{b})'\ P'\ P(Y - X\tilde{b})}{n-k-1}$

$$= \frac{(Y - X\tilde{b})'\ A^{-1}\ (Y - X\tilde{b})}{n-k-1} = \frac{\tilde{e}'\ A^{-1}\ \tilde{e}}{n-k-1}$$

This estimator of the variance is unbiased. Now by substituting $\tilde{\sigma}_\varepsilon^2$ into (6.16) we obtain the estimated variance-covariance matrix of the GLS estimators which is:

(6.19) $\quad V\ (\tilde{b}) = \tilde{\sigma}_\varepsilon^2\ (X'\ A^{-1}X)^{-1} = \tilde{\sigma}_\varepsilon^2\ (X'\ P'P\ X)^{-1}$

In summarizing generalized least squares procedures, let us focus on where this leaves us with respect to estimation. We need information about Φ as such or about A. We can use a combination of (6.15), (6.18) and (6.19) to obtain numerical values of the estimators. However, this requires a considerable amount of information which may not be available. In terms of the problem of serially correlated residuals the required information entails knowledge about ϕ or some outside estimate of it or, alternatively, knowledge of ρ or an estimate of it. If we estimate ρ by iterative procedures and then apply generalized least squares, which is equivalent to converting all data by the estimated autoregression coefficient and applying ordinary least squares.

Chapter 7 - Heteroscedasticity

One of the assumptions underlying the classical least squares linear regression model is that the variances of the disturbances are homogeneous. This is referred to as homoscedasticity. Homoscedasticity is embodied in $E(\varepsilon\varepsilon) = \sigma^2 I$. When this assumption is violated the variances are not homogeneous and this is referred to as heteroscedasticity.

There are at least two reasons why heteroscedasticity may occur. First, some of the observations being used in the regression analysis may be "less reliable" than others. This may occur because of the way the observations were generated or more likely by the way they were recorded. What this usually means is that the variances of the observations are not all equal. The covariances may still be zero if the disturbances are independent but:

(7.1)
$$E(\varepsilon\varepsilon') = \sigma^2 I \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \cdots & 0 \\ 0 & \sigma_2^2 & 0 & 0 \cdots & 0 \\ \cdot & & & \cdot & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 0 & 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

where σ_1^2, σ_2^2, ---, σ_n^2 are not all equal. Second, the heteroscedasticity may occur because for one reason or another disturbances aren't independent. In this case not only may the variances not all be equal but all the covariances will not be zero.

The consequences of applying ordinary least squares methods when heteroscedasticity exists are that the resulting estimators are unbiased but they will tend to be inefficient. Hence, tests such as the t and F and predictions will be inefficient.

Let us consider a heteroscedasticity situation where:

(7.2) $E(\varepsilon_t^2) = \sigma^2 C_t$ where $C_t \neq 0$

and

(7.3) $E(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$

The covariances are zero and the heterogeneity of variances may be expressed in some proportional terms. The variance-covariance matrix of the disturbances is:

(7.4) $E(\varepsilon\varepsilon') = \phi = \sigma^2 A = \sigma^2$

$$\begin{bmatrix} C_1 & 0 & 0 \cdots 0 \\ 0 & C_2 & 0 \cdots 0 \\ 0 & 0 & C_3 \cdots 0 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 0 & 0 & 0 \cdots C_n \end{bmatrix}$$

Now let us use the generalized least squares framework for estimation. We have:

(7.5) $A^{-1} =$

$$\begin{bmatrix} \frac{1}{C_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{C_2} & \cdots & 0 \\ 0 & 0 & \frac{1}{C_3} \cdots & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & \cdots & \cdots & \frac{1}{C_n} \end{bmatrix}$$

Letting

(7.6) $P =$

$$\begin{bmatrix} \frac{1}{\sqrt{C_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{C_2}} & \cdots & 0 \\ 0 & 0 & \frac{1}{\sqrt{C_3}} \cdots & 0 \\ \cdot & & & \cdot \\ 0 & \cdots & \cdots & \frac{1}{\sqrt{C_n}} \end{bmatrix}$$

then

$$P'P = P^2 = A^{-1} \tag{7.7}$$

Now let us premultiply both of

$$Y = X\beta + \varepsilon \tag{7.8}$$

by P and we obtain

$$PY = PX\beta + P\varepsilon \tag{7.9}$$

The variance-covariance matrix for the transformed disturbances is:

$$E[(P\varepsilon)(P\varepsilon)'] = E[P\varepsilon\varepsilon' P'] = \sigma^2 I. \tag{7.10}$$

We can apply ordinary least squares to (7.9). The least squares estimator is:

$$\tilde{b} = (X'P^2X)^{-1} X'P^2Y = (X'P'PX)^{-1} X'P'PY \tag{7.11}$$

with a variance-covariance matrix

$$V(\tilde{b}) = \sigma^2 (X'P^2X)^{-1} = \sigma^2 (X'P'PX)^{-1} \tag{7.12}$$

To obtain numerical estimates of (7.9) an estimate of σ^2 is required. This estimate may be obtained as follows:

$$\tilde{\sigma}^2 = (\tilde{Pe})' (\tilde{Pe}) = \frac{\tilde{e}' P' P\tilde{e}}{n-k-1} = \frac{(Y - X\tilde{b})' P'P (Y - X\tilde{b})}{n-k-1} = \frac{(Y - X\tilde{b})' A^{-1} (Y - X\tilde{b})}{n-k-1} \tag{7.13}$$

In order to adjust for the type of heterogeneous variances indicated in (7.1), values of the proportionality constants, C's, have to be obtained. There are three possible sources. First, is to assume values for the C's which is not very satisfactory if there is no prior information about magnitude of the variance. Second, prior information based on other studies which investigated the same phenomena but focused on the heterogeneous variance issue may be used. Third, if the observations are replicated, say five or so for each set of observed values of the exogenous variables, then the variances for each set of replications could be computed and hypotheses regarding homogeneity of variances tested. Unfortunately, for most types of problems with which economists are concerned

data with replications are not involved and it not possible to obtain estimates of the proportionality constants from the sample data.

Heteroscedasticity arises from serial correlation and is really a serial correlation problem. This is the type which we can do the most with. Heteroscedasticity arising from this source should be handled by focusing on the serial correlation problem and procedures applicable to serial correlation problems should be utilized. For example,

$$(7.14) \qquad \phi = \frac{\sigma_\delta^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & - & - & - & \rho^{n-1} \\ \rho & & & & & & \rho^{n-2} \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \rho^{n-1}\rho^{n-2} & & \rho^{n-3} & - & - & & 1 \end{bmatrix}$$

diagonals are 1 so σ_δ^2 is a common variance.

Finally, in most economic analyses utilizing multiple regression models it is not possible to verify the homoscedasticity assumption. Generally, information is not available to allow testing of hypotheses concerning the homogenity of variances. Hence, the assumption is made without any attempt at verification.

In subsequent chapters heteroscedasticity among equations in systems of equations will be considered. In such cases the variance-covariance matrix of variances and covariances of equations in the system can usually be estimated from the sample of data being used to estimate the parameters of the equations of the system. After the estimated variance-covariance matrix is obtained, GLS methods may be used to obtain more efficient estimators.

Chapter 8 - Errors in Variables

Introduction

There are three types of errors in variables: (1) errors in only the endogenous variables, (2) errors in only the exogenous variables (one or more), and (3) errors in both endogenous and exogenous variables. Each of these will be discussed briefly. [1/]

Errors Only in the Endogenous Variable

Suppose we are concerned with the relation:

(8.1) $Y = X \beta + \varepsilon$

where Y is the observed vector of the endogenous variable and the X matrix is measured without error. Further, suppose that the observed Y contains measurement errors of one sort or another such that

(8.2) $Y = Y_T + u$

where Y_T is the vector of the true values of the endogenous variable and u is the vector of measurement errors. Then the model which represents the relationship between the true values of the endogenous variable and the predetermined variable is:

(8.3) $Y_T = X \beta + (v - u)$

where the relevant disturbance term is $(v - u)$. It consists of two components, the true disturbance, v, and the measurement error, u. Thus ε in (8.1) really is

(8.4) $\varepsilon = v-u$

1/ Also see J. Johnston, Econometric Methods, 1972, pp. 281 - 291 and Jan Kmenta, Elements of Econometrics, 1971, pp. 307 - 322 for problems of errors in variables.

The effects of this type of measurement error in the endogenous variable depends on the characteristics of u. If u is independent of the predetermined variables and if

(8.5) $E(u) = 0$

the least squares method applied to (8.1) yields unbiased estimators of β but the estimators will not be efficient. The estimators of the regression parameters are:

$$(8.6) \quad b = (X'X)^{-1} X'Y = (X'X)^{-1} X' [X\beta + (v - u)]$$
$$= (X'X)^{-1} X'X\beta + (X'X)^{-1} X'v - (X'X)^{-1} X'u$$
$$= \beta + (X'X)^{-1} X'v - (X'X)^{-1} X'u$$

Now if both v and u are independent of the predetermined variables and have expected values of zero then,

$$(8.7) \quad E(b) = \beta + (X'X)^{-1} X' E(v) - (X'X)^{-1} X' E(u)$$
$$= \beta$$

Now

$$(8.8) \quad (b - \beta) = (X'X)^{-1} X' (X\beta + \varepsilon) - \beta$$
$$= (X'X)^{-1} X' [X\beta + (v - u)] - \beta$$
$$= \beta + (X'X)^{-1} X' (v-u) - \beta$$
$$= (X'X)^{-1} X' (v - u)$$

and

$$(8.9) \quad (b - \beta) (b - \beta)' = [(X'X)^{-1} X' (v - u)] [(X'X)^{-1} X' (v - u)]'$$
$$= (X'X)^{-1} X' (v - u) (v - u) (v - u)' (X'X)^{-1} X'$$

The variance and covariance of the estimators are:

$$(8.10) \quad V(b) = E[(b - \beta) (b - \beta)'] = E [(X'X)^{-1} X' (v - u) (v - u)'X(X'X)^{-1}]$$

but since elements of $(X'X)^{-1}$ and $X(X'X)^{-1}$ are taken as fixed:

$$(8.11) \quad V(b) = (X'X)^{-1} X' E[(v - u)(v - u)'] X(X'X)^{-1}$$
$$= (X'X)^{-1} X' E[vv' - uv' - vu' + uu'] X(X'X)^{-1}$$
$$= (\sigma_v^2 + \sigma_u^2 - 2\sigma_{vu}) (X'X)^{-1}$$

With the true disturbances, v, and the measurement errors, u, independent of each other the variance-covariance matrix of the estimators is:

$$(8.12) \quad V(b) = (\sigma_v^2 + \sigma_u^2) (X'X)^{-1}$$

and hence the variance of ε consists of

$$(8.13) \quad \sigma_\varepsilon^2 = \sigma_v^2 + \sigma_u^2$$

Note that if the expected value of the measurement error is not zero then the estimators are not unbiased.

If the residuals are serially correlated then $e = (\hat{v} - \hat{u})$ will exhibit some type of systemmatic pattern. The residuals are an estimate of the combined effect of the disturbance and the measurement error. The measurement error in this case is acting as on omitted variable which varies in a systemmatic pattern. This problem would be handled as a serial correlation problem. When measurement error in the endogenous variable is the source of the serial correlation in the residuals, incorporating some type of autoregressive scheme into the regression equation is a meaningful way of handling this type of measurement error problem.

Errors Only in the Predetermined Variables

Let us consider the case where the endogenous variable is measured without error but the predetermined variable(s) are subject to measurement error. Let:

$$(8.14) \quad X = X_T + w$$

where X_T is the matrix of the true values of the predetermined variables and w is the matrix of measurement errors in the predetermined variables. Now suppose the relationship of interest among the true values of the variables is:

$$(8.15) \qquad Y = X_T \beta + v$$

where v is the true disturbance and fulfills the conditions $E(v) = 0$ and $E(vv') = \sigma_v^2 I$

Now suppose we attempt to obtain the least squares estimators of β from the observed values of the predetermined variables which contain measurement errors. The least squares estimators of the coefficients are:

$$(8.16) \qquad b = (X'X)^{-1} X'Y = [(X_T + w)' (X_T + w)]^{-1} [(X_T + w)'Y]$$

$$= [(X_T + w)' (X_T + w)]^{-1} [(X_T + w)' (X_T \beta + v)]$$

$$= (X_T' X_T + X_T' w + w' X_T + w'w)^{-1} [(X_T' X_T + w' X_T) \beta + (X_T + w)'v]$$

and

$$(8.17) \qquad E(b) = [X_T'X_T + E(X_T'w) + E(w'X_T) + E(w'v)]^{-1} [X_T' X_T \beta + E(w'X_T) \beta + E(X_T' v) + E(w'v)]$$

v is assumed to be independent of X_T and w so

$$(8.18) \qquad E(b) = [X_T' X_T + E(X_T' w) + E(w'X_T) + E(w'w)]^{-1} [(X_T' X_T) \beta + E(W'X_T)\beta + 0 + 0]$$

$$= [X_T' X_T + E(X_T' w) + E(w'X_T) + E(w'w)]^{-1} [(X_T'X_T)\beta + E(w'X_T)\beta]$$

The measurement errors w may or may not be independent of the true values of the predetermined variables. The magnitudes of the measurement errors may depend on the magnitudes of the true values of the predetermined variables, directly or inversely. But suppose they are independent of

the true values of the predetermined variables then:

$$(8.19) \qquad E(b) = [X_T' X_T + E(w'w)]^{-1} (X_T' X_T) \beta$$

Clearly $E(b) \neq \beta$ and hence the least squares estimators are not unbiased. Furthermore, the bias will not disappear as sample size becomes very large i.e., approaches infinity so that the least squares estimators are inconsistent. It is not meaningful to consider the efficiency of biased and inconsistent estimators nor is it meaningful to perform test of significance.

If it is at all possible to estimate the magnitudes of the measurement errors corresponding adjustments may be made on the appropriate predetermined variables. With economic data this is usually not possible. If the measurement errors are proportional to the true values of the predetermined variables, i.e., large measurement errors associated with large values of X_T and small measurement errors associated with small values of X_T or vice-versa the result is likely to be heteroscadasticity. Adjusting for the heterogenous variances using generalized least squares may provide some relief but even then the basic underlying problem will remain.

The three methods mentioned in the next section for coping with measurement errors in both the endogenous and predetermined variables are also applicable if only the predetermined variables are subject to measurement errors.

Errors in Both Endogenous and Predetermined Variables

Let us consider the more usual base in practice and that is when both the endogenous and predetermined variables are subject to measurement errors. Let,

$$(8.20) \qquad Y = Y_T + u$$

and

$$(8.21) \qquad X = X_T + w$$

where Y and X are observed values of the endogenous variable and predetermined variables, respectively.

Y_T and X_T are the true values of the endogenous variable and predetermined variables, respectively.

u are the measurement errors of the endogenous variable.

w are the measurement errors of the predetermined variables (one or more predetermined variables).

Now suppose the relationship of interest among the true values of the variables is:

$$(8.22) \qquad Y_T = X_T \beta + v$$

where v is the true disturbance and fulfills the conditions $E(v) = 0$ and $E(vv') = \sigma_v^2 I$.

The least squares estimators based on the observed values of the endogenous and predetermined variables are:

$$(8.23) \qquad b = (X'X)^{-1} X'Y = [(X_T + w)' (X_T + w)]^{-1}[(X_T+w)' (X_T + u + v)]$$

$$= [(X_T + w)' (X_T + w)]^{-1}[(X_T' X_T + w'X_T)\beta + (X_T + w)'u + (X_T+w)'v]$$

$$= [(X_T' X_T) + X_T' w + w'X_T + w'w]^{-1} [X_T' X_T) \beta + (w'X_T)\beta + X_T' u + w'u + X_T' v + w'v]$$

and

$$(8.24) \qquad E(b) = [(X_T'X_T) + E(X_T'w) + E(w'X_T) + E(w'w)]^{-1} [(X_T'X_T) \beta + E(X_T'u) + E(w'u) + E(X_T'v) + E(w'v)].$$

v is assumed to be independent of X_T and w so

$$(8.25) \qquad E(b) = [(X_T' X_T) + E(X_T'w) + E(w'X_T) + E(w'w)]^{-1} [(X_T'X_T) \beta + E(w'X_T) \beta + E(X_T'u) + E(w'u)]$$

The measurement errors w and u may or may not be independent of the true values of the predetermined variables. As in the previous case the magnitudes

of the measurement errors may depend on the magnitudes of the true values of the predetermined variables, directly or inversely. If they are independent then $E(X_T' w) = E(w'X_T) = E(X_T' u) = 0$ and

$$(8.26) \qquad E(b) = [(X_T' X_T) + E(w'w)]^{-1} [(X_T' X_T)\beta + E(w'u)]$$

Furthermore, the measurement errors in the predetermined variables may not be independent of the measurement errors in the endogenous variable in which cases $E(w'u) \neq 0$. If w and u are independent then

$$(8.27) \qquad E(b) = [(X_T' X_T) + (w'w)]^{-1} (X_T' X_T) \beta$$

If both the endogenous and exogenous variables contain measurement errors, least squares methods yield estimators which are not only biased but they are inconsistent. To adjust for errors in both variables usually requires information which is not available. Furthermore, errors in variables are not easy to identify and the task is even more difficult when both the endogenous and exogenous variables contain measurement errors.

There are three common approaches to the problem of errors in both endogenous and exogenous variables. The classical approach is to utilize maximum likelihood estimators. In this approach the likelihood function is set up with all the relevant error terms and then the maximum likelihood estimators which contain the relevant error terms in addition to the appropriate variables are obtained. From a practical application viewpoint the difficulty with this approach is that it requires information about the errors which generally is not available.

A second approach is to utilize sub-groups of the data to obtain the estimates. Wald[2/] suggested ordering the exogenous variable in ascending order and dividing the data into two sub-samples. Then the slope coefficient

---

[2/] A. Wald, "The Fitting of Straight Lines if Both Variables are Subject to Errors", Annals of Mathematical Statistics, Vol. 11, 1940, pp. 284 - 300.

is estimated as the difference between the two groups in the means of the endogenous variables over the difference between the two groups in the means of the exogenous variable. The intercept is obtained in the usual way after the slope coefficient has been estimated. These estimators are consistent. Bartlett's [3/] method is similar to Wald's except he does not confine himself to using half the data in each sub-sample but in fact discards some of the data in the center of the array.

A third approach entails the use of instrumental variables. The instrumental variable technique involves choosing an exogenous variable which is not included in the model and using it as an operational instrument to circumvent the errors in variables problem. The instrumental variable selected is used to perform a transformation on the original data. The particular instrumental variable selected must be independent of the errors in both the endogenous and exogenous variables. Furthermore, the instrumental variable should be fairly highly correlated with the exogenous variable. A separate instrumental variable is needed for each exogenous variable containing measurement error. This method also yields a set of consistent estimators.

There are some difficulties in the application of the instrumental variable technique for errors in variables problems. The instrumental variables are choosen on an arbitrary basis and hence different instruments yield different estimates. Second, it is difficult and sometimes impossible to check whether the instrumental variable or variables are independent of the measurement errors in both endogenous and exogenous variables. Finally, consistency property is considered the ruling criterion and a consistent estimator may have undesireable small sample properties such as large sampling variances.

[3/] M. S. Bartlett, "The Fitting of Straight Lines if Both Variables are Subject to Error", Biometrics, Vol. 5, 1949, pp. 207 - 242.

Word of Caution About Measurement Errors

It is generally taken for granted that if measurement error exist that some effort should be made to make some adjustments for such errors and to take them into consideration in the estimation of the parameters of the model. However, whether any adjustments should be made depends on how the information is being used by the particular decision making units of interest. And most important is whether or not the particular decision making units base their decisions on the "true values" of variables and relationships between the "true values" of variables. If they make decisions based on "true values" of variables then measurement errors need to be accounted for and all the material discussed in earlier parts of this chapter are extremely important. However, this would imply that either the decision making units have access to data on "true values" of variables or they have some means of accounting for measurement errors. Both of these implications suggest means of accounting for measurement errors in estimation.

But what if the decision making units of interest have access only to the data that contains measurement errors and in fact make their decisions based on this data? Should the measurement errors be taken into consideration in the estimation of the parameters of the model? Suddenly the whole notion of what a measurement error is becomes all muddled. When dealing with most types of economic data should the whole concept of measurement error be relative to some ideal theoretical norm or relative to the decision making unit utilizing the data? If the detailed behavior of the mechanism generating the measurement error can be identified then clearly adjustments in estimation and adjustments for decision making can be made to satisfy all needs. But generally such details about the measurement errors can't be obtained. And if decisions are made on the basis of the data containing measurement errors then it would appear that these data containing measurement errors are the relevant ones for estimation. As far as the decision

making units are concerned there are no measurement errors or compensations are made for the measurement errors. Under these conditions applying ordinary least squares methods to the data containing measurement errors would seem appropriate.

It should be obvious that if the decision making units are basing their decisions on data containing measurement errors and on relationships between variables containing measurement errors, that adjustments for such measurement errors in estimation is in fact introducing an error into the decision making process. Hence, this is one good reason why concern about measurement errors has to be approached very carefully. This is also why much empirical estimation and analysis is performed with data that clearly contain some types of measurement errors yet the analyses generate information useful for decision making. If one is practical and realistic one would suspect that almost all types of secondary data contain some types of measurement errors. And furthermore, there are no obvious and clear cut guides that indicate exactly when and what kinds of adjustments for measurement errors should be made. Don't misinterpret this last statement as implying that measurement errors in variables really don't pose any kind of serious problems because they may and furthermore don't misinterpret the statement as an unqualified indicator to use "bad" data in estimation. Carefully assess the nature of measurement errors in variables and carefully appraise how the data and estimates will be used; then use good judgement with regard to whether any adjustments for measurement errors should and can be made and if so, what types of adjustments.

Chapter 9 - Multicollinearity

Introduction

One of the assumptions underlying the general linear model for the purpose of estimating its coefficients by least squares procedures is that the matrix of predetermined variables, X, has full column rank. Full column rank of this matrix means that the columns of X are linearly independent. This assumption is violated if there is linear dependency between two or more predetermined variables. The existence of linear dependency means that two or more of the predetermined variable are perfectly correlated. Under such conditions it is not possible to obtain $(X'X)^{-1}$ and the least squares procedure breaks down so that it is not possible to obtain estimates of the parometers of the linear model.

The extreme case of multicollinearity involving perfectly correlated predetermined variables is a rarity. The opposite extreme is the case where the degree of correlation among predetermined variables is zero. This means that there is no linear dependences among the predetermined variables or that the predetermined variables are orthozonal. This extreme case is also not very revelant unless the predetermined variables are being controlled experimentally and have been designed to be orthozonal.

The multicollinearity problems of greatest concern arise from situations where a degree of linear association exists between two or more predetermined variabled but not perfect linear association. The problem of multicollinearity is a problem of degree of linear association and thus correlation among predetermined variables. The severity of the impacts of multicollinearity depend on the degree of correlation among predetermined variables and the variability exhibited by the predetermined

variables. The higher the correlation among the predetermined variables the more severe the effect of multicollinearity.

Consequences of Multicollinearity

Multicollinearity may not be serious if $(X'X)^{-1}$ can be obtained and if focus is not on interpretation of individual regression parameters or the testing of hypotheses about individual regression parameters. When there is a high correlation among predetermined variables it is difficult if not impossible to isolate and interpret the separate effects of variables with any degree of confidence. Yet interpretation of the recession as a whole and predictions for sets of values of all predetermined variables may be made with little or no difficulty. Estimators of the β coefficients when obtainable are unbiased. Also, the estimator of the disturbance variance, σ^2, is not imparied.[1/] However, the variances of the estimators tend to be large and hence the estimators tend to be inefficient from the viewpoint of having large variances. These large variances of the estimators arise because of large diagonal elements of $(X'X)^{-1}$. As the correlation among predetermined variables increased the $(X'X)$ approaches singularity and the determinant of $(X'X)$ tends to approach zero which yields an unstable inverse matrix, $(X'X)^{-1}$.

When multicollinarity exists difficulties arise in testing hypotheses about individual regression parameters. The large variances of the estimators result in many of the conventional statistical tests such as the t and F being quite inefficient. Frequently conclusions that the testing statistic is not significant are made when in fact it may be.

1/ See the article by T. Haavelmo, "Remarks on Frisch's Confluence Analysis and Its Use in Econometrics," in Tjalling C. Koopmans (ed) Statistical Inference in Dynamic Economic Models, Cowles, Commission Monograph

Almost always the type of mistake one will make is calling some test non-significant at a chosen probability level when in fact it may have been.

Some difficulties also arise in interpreting the estimators. The estimated coefficients are generally intepreted as being the amount of change in the endogenous variable associated with a one unit change in the particular predetermined variable under consideration. When multicollinearity exists the changes associated with the correlated predetermined variables tend to be confounded. It is not possible to change a predetermined variable which is correlated with other predetermined variables without concurrently having the values of the other intercorrelated predetermined variables, also changing. Hence, when multicollinearity exists the estimated coefficient of a given predetermined variable is a sort of composite amount associated with that variable and all of the other correlated predetermined variables included in the equation.

Identification of Multicollinearity Problems

The extreme case where there is perfect positive or negative correlation between two or more predetermined variables is easy to identify. The simple correlation coefficients are either +1 or -1, depending upon whether the linear association is positive or negative, and the least squares breaks down and by conventional methods it is not possible to obtain an inverse of the (X'X) matrix.

Usually one should scrutinize the simple correlation coefficients between pairs of predetermined variables. High correlation coefficients, values approaching $\pm$ .90 or more, are signals suggesting that multicollinearity problems may arise.

The standard error and variances of the estimated coefficients are also useful indicators of multicollinearity. If Multicollinearity exists none or very few of the coefficients seem to be significantly different from zero or from some specified non-zero value. It is possible to have a rather high coefficient of determination, $R^2 > .90$ and based on the t-test to have judged none of the regression parameters to be significantly different from zero at commonly used levels of significance.

Alternative Approaches to Handling Multicollinearity Problems

There are several approaches which can be used to eliminate or circumvent the multicollinearity problem. No one approach is a solution for all situations. The method to be used needs to be selected in light of the particular situation being analyzed. Some possible approaches are discussed briefly.

1. Delete Variables

One of the first methods which comes to mind is to delete one or more of the highly correlated variables. In terms of explaining the variation in the endogenous variable, one may not be loosing much by the deletion. However, from an economic viewpoint, it may be quite critical to simply delete one or more of the exogenous variables. Even though we may not be loosing much statistical information by doing this we may be loosing considerable economic information. Recognize that by deleting one or more variables you are changing the original model. If the original model was theoretically sound, the model which you have after deletion of variables may not be. Furthermore, the model was developed on the basis of theoretical reasons but altered because of empirical and specifically sample data reasons.

Some information about the effects of the deleted variable may be acquired by obtaining the relationship between the deleted predetermined variable and the included predetermined variable with which it was highly

correlated. Given this relationship one can at least partly assess the effects of the deleted variable or translate the effects of the highly correlated included variable in terms of the deleted variable. One way of obtaining the relationship between the two predetermined variables of interest is to estimate as a separate side relationship the simple regression relation between the two variables as follows:

(9.1) $X_I = \alpha_0 + \alpha_1 X_D + U$

where

X_I is the included predetermined variable which is highly correlated with the deleted predetermined variable.

X_D is the deleted predetermined variable.

U is a random disturbance.

α_0 and α_1 are the regression parameters.

The corresponding estimating equation is:

(9.2) $X_I = a_0 + a_1 X_D + \hat{u}$

where a_1 gives the amount of change in the included predetermined variable associated with a one unit change in the deleted variable.

2. Combined Variables

One way of handling a high correlation between two predetermined variables is to combine the two variables and create a "new" variable which is included in the regression relationship. For example, if X_1 and X_2 are highly correlated then we may create a variable in the following manner.

(9.3) $Z = X_1 + k_1 X_2$ where $k_1 \neq 0$

In fact $k_1 = 0$ is simply another way of saying X_2 is deleted. If the value of k_1 is known from other than the sample data the problem can be handled simply and expediently. If a value of k_1 is not available then a_1 from (9.2) could serve as an approximation.

3. Prior Information and Restrictions

If prior information is available about the coefficients of one or more of the predetermined variables contributing to the mutlicollinearity problem, then this information may be incorporated into the model and taken into consideration in estimation. One source of such prior information are other studies. If estimates of one or more of the coefficients of the variables of concern are available and one feels comfortable with them, the parameters of these same variables in the relationship under consideration may be constrained to take on the values provided by the prior estimates. This has been done in demand studies where for example an income coefficient or income elasticity based on cross-sectional data was incorporated into a demand relationship concerned with a time series relation.

4. Transform Data

Sometimes transforming the data by a trigometric or logarithmic transformation will reduce the multicollinearity sufficiently to at least permit obtaining estimates of the coefficients. Similarly measuring the variables as deviations from their respective means will also reduce the multicollinearity enough to permit estimates of the coefficients to be obtained. Measuring the variables as deviations from their respective means tends to work well with polynomial functions whose terms include powers of a given variable, i.e., quadratic forms X and X^2, cubic functions, X, X^2 and X^3, etc.

One thing to recognize in performing transformations on the data is the consequences of such a transformation on the form of the relationship being estimated. If one is not careful, one can perform a transformation on the data which will reduce or eliminate the multicollinearity but one may end up with a relationship and/or forms of variables which cannot be meaningfully interpreted with respect to the problem being analyzed.

5. Correct The Inverse

In the case where the inverse of the moment matrix and estimates of the coefficients are obtainable, much of the efficiency difficulties stem from the unstableness of the inverse. Diagonal elements of the inverse are large and $(X'X)^{-1}(X'X)$ and $(X'X)(X'X)^{-1}$ deviate from the identity matrix of appropriate size. One way of reducing the difficulrues arising from the "mechanical aspects" of the problem is to adjust the inverse so as to make it more similar to the true inverse. For a true inverse we would have exactly:

$$(9.4) \quad (X'X)(X'X)^{-1} = I = (X'X)^{-1}(X'X).$$

Let

$$(9.5) \quad A = (X'X)$$

and

$$(9.6) \quad C = (X'X)^{-1}$$

then we can compute AC and the matrix $(2I - A\,C).(2I - AC)$ is the same as AC except the diagonal elements are subtracted from 2 and the signs of all other off-diagonal elements are reversed.

Now we can compute

$$(9.7) \quad C'' = C(2I - AC)$$

where C" is a corrected inverse. We can check this inverse to see if the inverse has been improved adequately where there is no hard fast rule for specifying what is adequate. If

$$(9.8) \quad C''A$$

still deviates from the identity more than what is felt to be desirable the interative steps are repeated. Next compute the matrix

$$(9.9) \quad C''' = C''(2I - AC'')$$

C"' is the new adjusted inverse. Again check it against A to see if C"'A is sufficiently close to the identity matrix. If one is satisfied with how close it is one can stop. If not, repeat the steps once again

and obtain C''''. One can continue the interative procedure until a desirable approximation of the inverse, $(X'X)^{-1}$ is obtained.

An important thing to recognize in this procedure is that it focuses strictly on the mechanics of the inverse. It does nothing about the basic cause, namely, the high correlation between two predetermined variables or among pairs of predetermined variables. This method focuses strictly on improving the accuracy of the inverse given that there are problems with the inverse which arise because there is multicollinearity.

6. Simultaneous Model

One thing which also needs to be considered when multicollinearity problems arise is whether the model being considered is incorrect because it is only a part of a larger model. The high correlation between two or among more than two ot the predetermined variables may be due to the fact that predetermined variables are not really predetermined but they are jointly determined. If this is the case, this suggests there are more relationships in the model and the one under consideration is only one of several. If simultaneity is really the basic problem, then the model needs to be reformulated as a simultaneous system of equations.

If the correlation among predetermined variables is in fact simply a linear association among variables that truly are predetermined, then formulating a simultaneous equation model is not appropriate. However, if variables are correlated because they are jointly determined and there is multidirectional causality among them, then the simultaneous system is appropriate. Often economic variables such as prices and quantities of substitutes and complements are highly correlated because the economic system is operating like our economic theory says they should and their values are jointly determined.

Chapter 10 - Lagged Variables

Introduction

Lagged variables include both lagged exogenous and lagged endogenous variables. Both current and lagged values of exogenous variables may be included in a linear multiple regression if the behavior of the economic units being analyzed suggests that such a specification is appropriate. Specific forms of lags in exogenous variables may essentially result in relationships that are equivalent to those obtained by incorporating a lagged endogenous variable in the model.

There has been considerable interest in distributed lag models. Irving Fisher[1/], Alt[2/], and Tinbergen[3/] were some of the earlier users of lagged variables. Koyck[4/] utilized distributed lags to analyze investments and Nerlove[5/] championed the notion of a dynamic partial adjustment model for analyzing producer and consumer behavior. More recently Shirley Almon[6/]

1/ Irving Fisher, "Our Unstable Dollar and the So-Called Business Cycle", JASA, Vol. 20, 1925, pp. 179-202 and "Note on a Short-Cut Method for Calculating Distributed Lags", Bulletin of the International Statistical Institute, Vol. 29, 1937, pp. 323-327.

2/ F. F. Alt, "Distributed Lags", Econometrica, Vol. 10, 1942, pp. 113-128.

3/ J. Tinbergen, "Long-Term Foreign Trade Elasticities, Metroeconomica, Vol I, 1949, pp. 174-185.

4/ L. M. Koyck, Distributed Lags and Investment Analysis, North-Holland Publishing Co. Amsterdam, 1954.

5/ Marc Nerlove, "Distributed Lags and Estimation of Long-Run Supply and Demand Elasticities", Theoretical Considerations, Journal of Farm Economics, Vol. XL, No. 2, May 1958, pp. 301-311.

6/ Shirley Almon, "The Distributed Lag Between Appropriations and Expenditures", Econometrica, Vol. 33, No. 1, January 1965, pp. 178-196.

utilized polynomial distributed lags to analyze capital appropriations and expenditures. These are just a few pieces of literature concerned with lagged variables and lag distributions. They do reflect rather wide differences in approaches and developments in utilizing lagged variables in economic analyses.

Generally a discussion of lagged variables starts with lagged exogenous variables. Here we will start with the lagged endogenous variable and the partial adjustment and expectation models that are sometime referred to as the "Nerlovian distributed lags". Then some limited attention will be devoted to lags in exogenous variables and their linkage to models containing the lagged endogenous variable. Polynomial lags will be one of the last topics to be discussed in this chapter.

Underlying Motivation

Distributed lags arise in theory when economic stimuli produce their effects only after some lag in time, so that the effects are not felt all at once, at a single point of time, but are distributed over a period of time. For example, a change in the quantity demanded of a good in response to a price change or an income change would not be felt immediately but the effects are felt gradually so that the full effect is realized only after a period of time has elapsed. Sometimes this period of time may be considerable. Just how considerable it may be depends on the kinds of rigidities and "stickiness" which may exist in the economy or sector being considered.

Lagged responses or adjustments are generally attributable to one or more of three factors. First, the lags may be due to psychological factors which include forces of habit and assumptions on the part of the consumer that changes are temporary and thus there is no need to attempt to fully adjust to the stimulus. Second, the lags may be due to technical factors

which, for example, may include such things as the inability to increase the use of frozen foods without first acquiring adequate freezer space. Third, the lags may be due to institutional factors which may include situations in which certain contractual items of expenditure or savings may need to be adjusted before shifts can be made in consumption patterns or the fact that some markets, particularly for durable goods, are imperfect in an economic sense.

The same notions that underlie distributed lag concepts also underlie the distinction between short and long-run elasticities concepts. Thus as Nerlove [7] puts it "the formulation of economic relationships containing distributed lags is related to the problem of formulating meaningful relationships among variables we can observe, and the problem of estimating distributions of lags is really the problem of estimating long-run elasticities.

There are at least three approaches to estimating a distribution of lag. First, no assumption need be made as to the form of the distribution. This was the approach taken by Alt[8] and Tinbergen[9]. The weights associated with various lags were considered an empirical issue. In other words, incorporate all conceivable lags into the model and estimate the relevant weights from the same sample of data. Because of the finite length of and serial correlation in most economic time series, this approach may not always be feasible. Second, a general form for the distribution of the lag may be assumed and parameters which define the exact ditribution estimated. The second approach is the one taken by Irving Fisher[10] and Koyck[11]. The

---

[7] Op. cit., p. 307.

[8] Op. cit.

[9] Op. cit.

[10] Op. cit.

[11] Op. cit.

assumption concerning the distribution of lag can't help but be arbitrary. Finally, an explicit dynamic model of behavior which implies a distributed lag only incidently may be developed. This approach is the one championed by Nerlove[12/] and leads to a direct interpretation of the distribution of lag in terms of economic behavior and in terms of the difference between short and long run elasticities. This is the form of lag that will be discussed first.

Nerlovian Lags

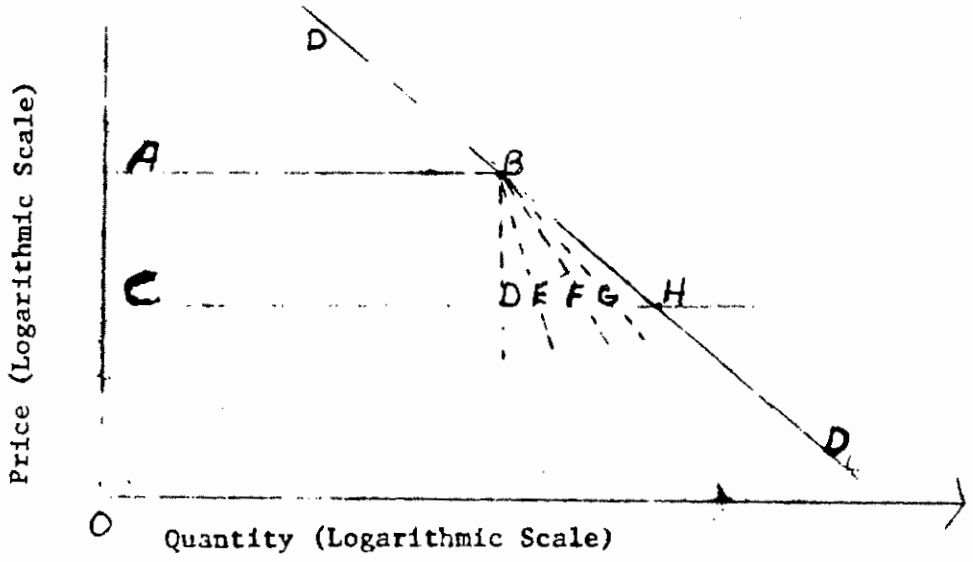
Distributed lag models have been used in supply and demand analysis as a means of estimating long-run elasticities. In agricultural economics in particular there has always been a concern about the long-run supply and demand relations for various agricultural products. The difficulty with obtaining empirical coefficients of any long-run economic relation stems from the fact that long-run equilibrium values of economic variables are not observable. To overcome this difficulty, assumptions are usually made about the relationships between observed values of variables and the long-run unobservable values of the variables. Such assumptions have led to various distributed lags.

Nerlove argues that short-run elasticities cannot be estimated because through each point on a long-run supply or demand relation there passes a "fan" of short-run curves, each one appropriate to a different interval of time. Therefore, the elasticity estimates of supply and demand relations are not unique in the short-run. The short-run elasticities differ depending on the initial position and the length of time allowed for adjustment. This is elementary but crucial to the estimation of elasticities.

[12/] Op. cit.

Let us illustrate this graphically using a demand situation for illustration purposes. To facilitate the illustration let's make the following assumptions: (1) Supply is perfectly elastic, so that the problem may be dealt with in the absence of difficulties caused by the existence of simultaneous relations among variables, (2) Observations on price and quantity are taken at discrete time intervals of a fixed length and full adjustment to changed conditions does not take place within the time length of the intervals, (3) A once and for all change in price is assumed, and (4) Assume a long-run curve D. Now in Figure 10.1 consider a price OA which is given by a supply-demand equilibrium. At this price, AB will be taken. Now suppose a shift in supply results in a new price, OC. It immediately changes from AB to CH. Instead

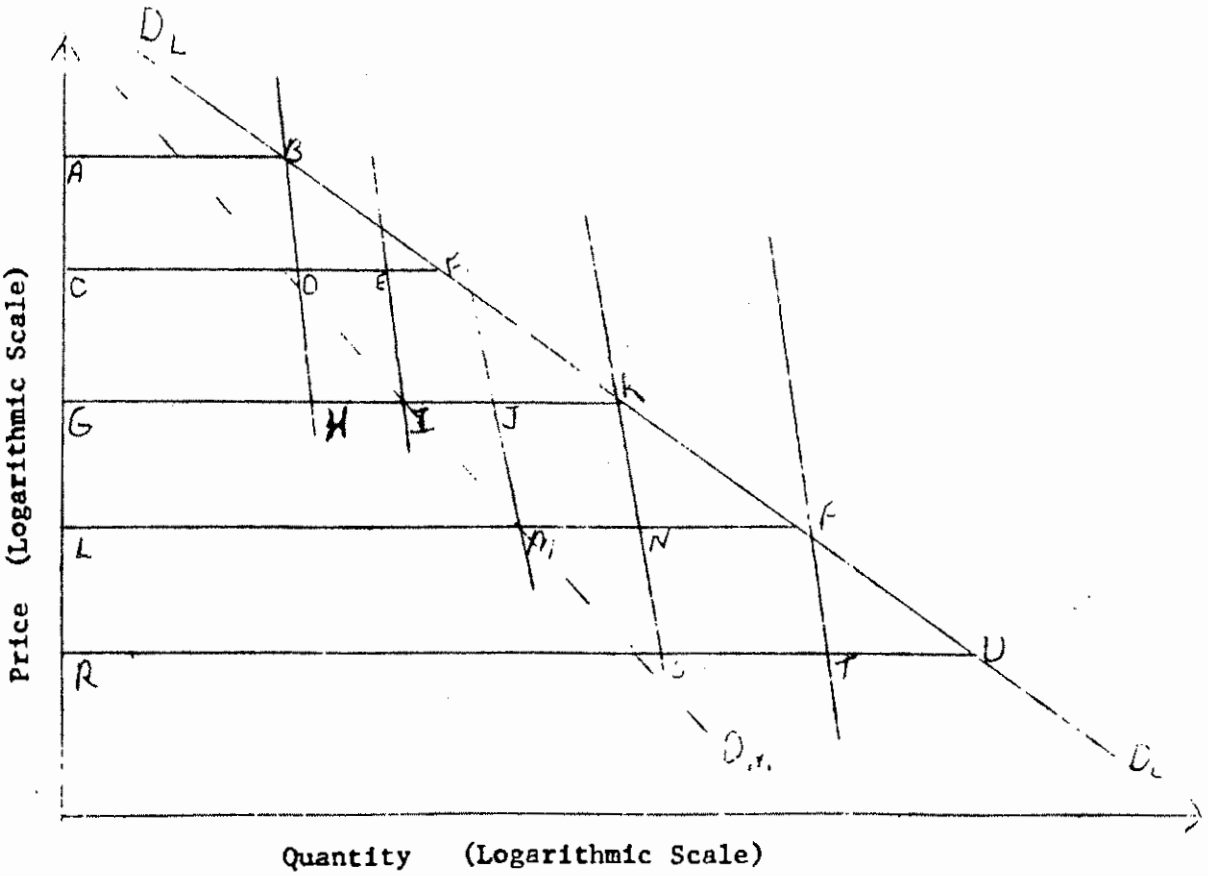
FIGURE 10.1 Adjustment of the Quantity Demand to a Once-and-For All Change in Price.



quantity demanded is determined by short-run demands BD, BE, BF, etc. Each of these represents different lengths of short-run. The quantity demanded associated with each of these short-runs is CD, CE, CF, etc. As time passes and the price remains at P_2 the quantity demanded approaches CH where H is a point on the long-run demand curve.

But in the real world prices don't simply exhibit a once-and-forever change. Realistically prices will be changing constantly and observed points never lie on the long-run demand curve. Let's attempt to illustrate this is Figure 10.2. Start with an initial equilibrium point at B on the long-run demand curve $D_L D_L$ associated with the price OA. Let the supply curve shift such that price decreases to OC, then to OG, to OL and finally to OR. As price changes from OA to OC consumers adjust the quantity demanded from AB to CD. If the price remains at OC, the consumer would demand CE the following period. And of course if the price remained there indefinitely they would eventually demand CF where F lies on the long-run demand curve. But price doesn't remain at OC indefinitely but assume that the next period it changes to OG. Consequently consumers move along a new short-run demand curve through point E to point I. Consumers demand slightly more at price OG than if the price had remained at OC. If price

FIGURE 10.2 Adjustment of the Quantity Demanded to Successive Changes in Price.



continued to decline a series of points D, I, M etc. would be observed all of which lie on different short-run demand curves passing through different points on the long-run demand curve. A curve passing through these points $D_M D_M$ is sort of a "mongrel curve" that has neither the average elasticity of the short-run curves nor the elasticity of the long-run curve. $D_M D_M$ is the curve that would be estimated if the whole problem of short and long run demand is neglected and is the demand curve which usually has been estimated. The position, elasticity, and shape of the estimated demand curve, $D_M D_M$, depends on the shape and positions of the short-run demand curves and the shifts in the supply curve and associated pattern of price changes. In fact it is questionable whether $D_M D_M$ is any kind of a demand curve and under the conditions specified here, it is really a "predictive curve" of supply and demand intersections but does not represent the demand structure. This is the crux of Nerlove's argument and is at the heart of his lag mechanism.

Expectations

Up to this point no mention has been made of uncertainty and expectations. Yet the usual distinction between long and short-run elasticities rests on the assumption that expectations are static, i.e., that people believe that current prices and incomes will persist indefinitely. If expectations are not static then the magnitudes of long-run elasticities need not exceed the magnitudes of short-run elasticities.

There are two serious problems that arise in analysis of non-static expectations. First, expectations of this type are generally not single valued and thus there is the issue of how to incorporate these in an analysis in a meaningful way and at the same time have the procedure be

simple and manageable. Second, usually there is a considerable number of non-quantifiable factors that influence the formation of expectations and explicitly incorporating these into the analysis is difficult.

One concept of expectations which treats them as single valued is the notion of an expected "normal" value (price, income, or other variable), which is the level about which future values of the variables are expected to fluctuate. If changes in "expected" normal values are induced by changes in current values of the corresponding variables the same simple models of expectations formulation may be constructed. For example, let Z_t^* be the expected "normal" level of a variable at time t, and let Z_t be the actual level. Then

$$(10.1) \qquad Z_t^* - Z_{t-1}^* = \gamma (Z_t - Z_{t-1}^*) \ , \quad 0 < \gamma \leq 1$$

where γ is called the coefficient of expectation if Z* and Z are expressed in actual variates or it is the elasticity of expectation if Z* and Z are expressed in logarithms. Note that (10.1) states that in each period the notion of what is normal is revised in proportion to the difference between what actually happened and what was previously considered "normal". When the elasticity or coefficient of expectations is one, expectations are static; when it is less than one the expected "normal" fluctuates less than do the actual values of the variable.

Note if supply depends on expected "normal" prices and demand on expected "normal" income and prices as well as on current prices, the introduction of non-static expectations in the form specified in (10.1) leads to another reason why long-run elasticities should exceed short-run elasticities. For example, if a consumer experiences a change in income he may not believe that the change will persist so there is no incentive to adjust fully to the change. But if this new level persists

through time, the consumer will believe in its permanence and respond accordingly. In other words there will be a tendency to adjust toward the long-run equilibrium quantity.

A Partial Adjustment Model

Let us now look at a simple dynamic model which leads to a distributed lag and which represents the phenomena reflected in Figure 10.1. Let q_t^* be the quantity demanded in the long-run equilibrium and let it be a function of current relative price, current real income, and other variables. Let q_t be the current quantity demanded. Now in the absence of changes in relative price real income or the other variables upon which quantity demanded depends, the current quantity demanded would change in proportion to the difference between the long-run equilibrium and the current quantity. Thus, this could be expressed as the following difference equation:

$$(10.2) \qquad q_t - q_{t-1} = \gamma(q_t^* - q_{t-1}) \qquad 0 < \gamma \leq 1$$

where γ is the constant of proportionality which is either the elasticity or coefficient of adjustment depending on whether the quantity is expressed in logarithms or actual variates.

Now let P_t be the logarithm of relative price and Y_t be the logarithm of real income. One representation of the long-run demand curve, $D_L D_L$ in Figure (10.1) is:

$$(10.3) \qquad q_t^* = a_0 + a_1 P_t + a_2 Y_t$$

where a_1 is the long-run price elasticity and a_2 is the long-run income elasticity. A disturbance term is not included in (10.3) at this time because it is not being viewed as a stochastic relation.

Now from (10.2) we see that

$$(10.4) \qquad q_t^* = \frac{q_t - (1 - \gamma)\, q_{t-1}}{\gamma}$$

Substituting into (10.3) yields

$$(10.5) \qquad \frac{q_t - (1 - \gamma)\, q_{t-1}}{\gamma} = a_0 + a_1 P_t + a_2 Y_t$$

and

$$(10.6) \qquad q_t = \gamma a_0 + \gamma a_1 P_t + \gamma a_2 Y_t + (1 - \gamma)\, q_{t-1}$$

Reformulating (10.6) in terms of a stochastic relation yields:

$$(10.7) \qquad q_t = \beta_0 + \beta_1 P_t + \beta_2 Y_t + \beta_3\, q_{t-1} + \varepsilon_t$$

where

$$\beta_0 = \gamma a_0$$
$$\beta_1 = \gamma a_1$$
$$\beta_2 = \gamma a_2$$
$$\beta_3 = 1 - \gamma$$

Hence, $\gamma = 1 - \beta_3$.

Now if we estimate (10.7) by ordinary least squares methods, the estimating equation is:

$$(10.8) \qquad q_t = b_0 + b_1 P_t + b_2 Y_t + b_3 q_{t-1} + e_t$$

where b_0, b_1, b_2, and b_3 are estimators of β_0, β_1, β_2, and β_3 respectively and e_t is the residual for time period t. The coefficient of adjustment (or elasticity of adjustment if the variables are in logarithms) is estimated by:

$$(10.9) \qquad \hat{\gamma} = 1 - b_3$$

where b_3 is the estimated coefficient of the lagged endogenous variable. The coefficient or elasticity of adjustment is always one minus the coefficient of the lagged endogenous variable.

The coefficients (a's) in (10.3) are long-run coefficients. If the variables are in logarithms the coefficients are long-run elasticities and if the variables are in actual values - then a_0 is the long-run intercept and the other a's are long-run slope coefficients. Note that the β's in (10.7) are some type of short-run parameters where the length of run pertains to the time interval represented by the t subscript. Also, note that the β's are products of the coefficient (or elasticity) of adjustment and the long-run coefficients. From (10.8) estimates of the β's can be obtained and from (10.9) an estimate of the coefficient or elasticity of adjustment can be obtained. Thus by dividing the b coefficients, except b_0, of (10.8) by the $\hat{\gamma}$ obtained in (10.9) the long-run estimates of the parameters of (10.3) are obtained. If (10.3) and (10.8) are expressed in logarithms the long-run price elasticity is $\frac{b_1}{\hat{\gamma}}$ and the long-run income elasticity is $\frac{b_2}{\hat{\gamma}}$. The long-run intercept is $\frac{b_0}{\hat{\gamma}}$. If the variables in(10.3) and (10.8) were expressed in actual terms $\frac{b_1}{\hat{\gamma}}$ and $\frac{b_2}{\hat{\gamma}}$ are estimates of long-run price and income slope coefficients respectively and $\frac{b_0}{\hat{\gamma}}$ is an estimate of the long-run intercept for the actual variate demand equation.

The γ coefficient is a proportionality constant and indicates the proportion of adjustment made toward the long-run equilibrium in one time period as measured by the time interval denoted by t when all other variables are non-variant. Hence, a onee-and-forever shock is being considered and all other variables remain constant. Thus, in the first time period γ percent adjustment toward long-run equilibrium is made. In a subsequent second time period $\gamma(1 - \gamma)$ additional adjustment is made or the total adjustment for two periods is $[\gamma + \gamma(1 - \gamma)]$. After the second time period $\{1 - [\gamma + \gamma(1 - \gamma)]\}$ adjustment toward long-run equilibrium remains and thus the proportion of adjustment made in the third period is

$\gamma\{1 - [\gamma + \gamma(1 - \gamma)]\}$. The process can be continued and the number of time periods required to achieve a specified proportion of adjustment toward long-run equilibrium can be determined. Note that if the model is in actual variates the proportion of adjustment refers to actual values of the variables and if the model is in logarithmic form the proportion of adjustment refers to logarithmic values of the variables.

To further illustrate the interpretation of the coefficient of adjustment let us consider an actual variate model with $\hat{\gamma} = .40$. In one time period .40 or 40 percent of adjustment toward long-run equilibrium is made and 160 or 60 percent remains. All other variables are assumed non-variant. In the second time period 40 percent of the remaining 60 percent of adjustment or 24 percent more is made. In two time periods, 64 percent of the adjustment is made. 40 percent of the remaining 36 percent or 14.4 percent is made in the third period. Thus, after three time periods 78.4 percent of adjustment toward long-run equilibrium would be accomplished.

Adaptive Expectation Model

In (10.1) a simple expectation model was presented but a slightly different version of expectation was used in developing the partial adjustment model. Underlying the partial expectation model in the notion that the difference in the value of an endogenous variable between time period t and t-1 is some proportion of the difference between the long-run equilibrium value in t and the observed value of the endogenous variable in t-1. The variable of interest in the simple demand relation illustration was the quantity demanded. Once again a simple demand relation will be considered but more direct use will be made of the expectation presented in (10.1).

Let us consider an economic relationship in which the value of the endogenous variable in time period t depends on an expectation of a predetermined variable formed in time period t rather than on an observed value of the predetermined variable in time period t. Drawing once again on a simple demand relationship for illustrative purposes we have:

$$(10.10) \quad q_t = \alpha_0 + \alpha_1 P_t^* + \alpha_2 Y_t$$

where P_t^* is a price expectation formed in time period t, Y_t is some relevant measure of income in period t, and q_t is the quantity demanded in time period t. The α's are parameters. To keep things simple as possible let us not include a random disturbance in (10.10).

In time period t-1 the quantity demanded is:

$$(10.11) \quad q_{t-1} = \alpha_0 + \alpha_1 P_{t-1}^* + \alpha_2 Y_{t-1}$$

where the parameters and variables are the same as in (10.10) except of course the variables are all lagged one time period. (10.11) may be solved for P_{t-1}^* which yields:

$$(10.12) \quad P^*_{t-1} = \frac{q_{t-1}}{\alpha_1} - \frac{\alpha_2}{\alpha_1} Y_{t-1} - \frac{\alpha_0}{\alpha_1}$$

Now looking at the expectation model in (10.1) expressed in terms of the price variable is:

$$(10.13) \quad P^*_t - P^*_{t-1} = \gamma (P_t - P^*_{t-1}) \quad , \quad 0 < \gamma \leq 1$$

and

$$(10.14) \quad P^*_t = \gamma (P_t - P^*_{t-1}) + P^*_{t-1}$$

$$= \gamma P_t - \gamma P^*_{t-1} + P^*_{t-1}$$

$$= \gamma P_t + (1 - \gamma) P^*_{t-1}$$

Substituting (10.14) into (10.10) leads to:

$$(10.15) \quad q_t = \alpha_0 + \alpha_1[\gamma P_t + (1 - \gamma) P^*_{t-1}] + \alpha_2 Y_t$$

$$= \alpha_0 + \alpha_1 \gamma P_t + \alpha_1 (1 - \gamma) P^*_{t-1} + \alpha_2 Y_t$$

and substituting (10.12) into (10.15) yields:

$$(10.16) \quad q_t = \alpha_0 + \alpha_1 \gamma P_t + \alpha_1 (1-\gamma) \left[\frac{q_{t-1}}{\alpha_1} - \frac{\alpha_2}{\alpha_1} Y_{t-1} - \frac{\alpha_0}{\alpha_1} \right] + \alpha_2 Y_t$$

$$= \gamma\alpha_0 + \gamma\alpha_1 P_t + \alpha_2 (Y_t - Y_{t-1}) + \gamma\alpha_2 Y_{t-1} + (1-\gamma) q_{t-1}$$

$$= \gamma\alpha_0 + \gamma\alpha_1 P_t + \alpha_2 \Delta Y_t + \gamma\alpha_2 Y_{t-1} + (1-\gamma) q_{t-1}$$

Note that the model contains both a first difference of the exogenous variable or variables and also the exogenous variable or variables lagged.[13/]

[13/] If the demand model were $q_t = \alpha_0 + \alpha_1 P^*_t$ and $P^*_t - P^*_{t-1} = \gamma (P_t - P^*_{t-1})$

then $q_{t-1} = \alpha_0 + \alpha_1 P^*_{t-1}$ and $q_t = \gamma\alpha_0 + \gamma\alpha_1 P_t + (1-\gamma) q_{t-1}$

The first difference and lagged exogenous variables don't appear. This is the case generally used for illustration.

Reformulating (10.16) in terms of a stochastic relation yields.

$$(10.17)\quad q_t = \beta_0 + \beta_1 P_t + \beta_2 \Delta Y_t + \beta_3 Y_{t-1} + \beta_4 q_{t-1} + \varepsilon_t$$

where $\beta_0 = \gamma\alpha_0$

$\beta_1 = \gamma\alpha_1$

$\beta_2 = \alpha_2$

$\beta_3 = \gamma\alpha_2$

$\beta_4 = 1 - \gamma$

Hence, $\gamma = 1 - \beta_4$ which is the way the coefficient of expectation is estimated. In this case the coefficient of expectation is a coefficient of price expectation. If the variables are in logarithms then γ is an elasticity of expectation. In either case γ indicates the proportion of the difference between the current observed price and last periods expected normal price is of the difference between the current period's and last period's expected normal price. As a coefficient of expectation the proportionality constant refers to variables measured as actual variates while as an elasticity of adjustment the proportionality constant refers to variables measured as logarithms.

Some Potential Estimation Problems

To estimate the parameters of (10.17) we would be concerned with the following estimating equation:

$$(10.18)\quad q_t = b_0 + b_1 P_t + b_2 \Delta Y_t + b_3 Y_{t-1} + b_4 q_{t-1} + e_t$$

where the b's are estimators of the β's and e_t is the residual for time period t and is an estimate of the disturbance ε_t. Some severe multicollinearity problems could arise in the estimation of (10.18) given that ΔY_t, Y_{t-1} and q_{t-1} all appear in the model as predetermined variables. Since multicollinearity is a sample phenomena whether or not multicollinearity problems arise will depend on the specific sample values of the

predetermined variables.

A potentially severe problem that may arise anytime the lagged dependent variable is incorporated into a linear model as a predetermined variable is the biasedness of the coefficients. In ordinary least squares for

(10.19) $F(b) = \beta$

to hold the covariance between the disturbance term and all predetermined variables must have expected values of zero. This will occur if there is no correlation between the disturbance term and each of the predetermined variables and

(10.20) $E(\varepsilon) = 0$

We know that in the demand relations considered that

(10.21) $F(q_t, \varepsilon_t) \neq 0$

and if the model is specified properly,

(10.22) $E(q_{t-1}, \varepsilon_t) \neq 0$

If (10.22) is true and letting Z represent the matrix of predetermined variables, then

(10.23) $E(b) = \beta + E(Z'\varepsilon)$

and

(10.24) $E(Z'\varepsilon) \neq 0$

because the expected covariance between the distrubance and at least one Z, namely q_{t-1}, is not equal to zero.

All the estimators will tend to be biased but with these types of models we are particularly concerned about the coefficient of the lagged endogenous variable. It is either a coefficient (or elasticity if variables are in logarithms) of expectation or a coefficient (or elasticity if variables are in logarithms) of adjustment used in estimating long-run coefficients. The bias in the coefficient of the lagged endogenous variable tends

to be upward. In other words a lagged endogenous variable used as a predetermined variable tends to account for "more variation in the endogenous then it ought to". It tends to "over account" for time related patterns in variables. An over-estimation of the coefficient of the lagged endogenous variable leads to an under-estimation of the coefficient (or elasticity) of expectation or adjustment and subsequently to an over-estimation of long-run coefficients and the time required for adjustment. Trend removal or accounting for trend in some manner may help reduce the over-estimation problem of the coefficient of the lagged endogenous variable. Inclusion of time measured in ascending or descending order i.e. 1,2,---,n or n, n-1, - - -, 1 is one way of attempting to account for at least part of the time related pattern. Trend removal or incorporating time in some form in the model does not alleviate the biasedness problem as such but it does tend to reduce the over-estimation problem concerning the coefficient of the lagged endogenous variable.

Koyck Types of Lags

The most general form of a model which allows for a distributed lag is as follows:

$$(10.25)\quad Y_t = \beta X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + - - - + \varepsilon_t$$

where

$$(10.26)\quad E(\varepsilon_t^2) = \sigma^2 \ , \ E(\varepsilon_s, \varepsilon_t) = 0 \text{ for } s \neq t$$

and X and ε are independent.

Although (10.25) is a more general than the formulations of a simple period lag in endogenous variable that we have considered thus far, its estimation presents some difficulty. The sample sizes with which we deal are finite so the infinite set of lagged variables have to be arbitrarily cut off at some point. Even then, there is likely to be a multicollinearity because successively lagged variables will tend to be highly correlated.

Some sort of restriction is in order and a natural one is to specify a priori that the coefficients of the successively lagged X's decline systemmatically as we go further back in time. In other words the closest lags carry the heaviest weight. This was originally suggested by Irving Fisher and later by Koyck. Koyck's specification is that the coefficients of (10.25) decline geometrically. The general expression for any coefficient of (10.25) is:

$$\beta_k = \beta\lambda^k \quad (k = 0, 1, 2, - - -) \tag{10.27}$$

where $0 < \lambda < 1$. Thus (10.25) may be re-written as:

$$Y_t = \beta X_t + \lambda\beta X_{t-1} + \lambda^2\beta X_{t-2} + - - - + \varepsilon_t \tag{10.28}$$

Lagging (10.28) by one period and multiplying through by λ yields

$$\lambda Y_{t-1} = \lambda\beta X_{t-1} + \lambda^2\beta X_{t-2} + \lambda^3\beta X_{t-3} + - - - + \lambda\varepsilon_{t-1} \tag{10.29}$$

Subtracting (10.29) from (10.28) gives

$$Y_t = \beta X_t + \lambda Y_{t-1} + \varepsilon_t + \lambda\varepsilon_{t-1} \tag{10.30}$$

which is similar to (10.7) except for the underlying structure of the disturbance term.

The particular formulation presented in (10.25) contains only one exogenous variable but different lags of that one exogenous variable. The framework is applicable to formulations containing more than one exogenous variable. Furthermore, all exogenous variables do not have to be lagged the same number of periods provided the phenomena being analyzed merits such consideration. Also, more than one exogenous variable may be lagged the same time periods considered in the lagging, but different weights given to different exogenous variables lagged the same number of periods.

Martin[14/] considered a nonlinear distributed lag model containing two lag parameters, one for price and one for income in an analysis of demand. The model assumes that the effects of all price and income changes diminish through time at constant but different geometric rates. Thus, it is assumed that income posses a different lag distribution from that of price.

Caution About Using Distributed Lags

Griliches[15/] presents a detailed survey of characteristics and estimation problems associated with geometrically declining distributed lag models. He offers a few rules to follow for anyone using distributed lag models.

1. If one is working with strongly trending data one should investigate whether the exogenous variables provide an adequate explanation for these trends. Do not throw the problem into the residual category without doing something about trend removal. The standard statistical theory applies only to the case of stationary disturbances. In practice, with estimated roots close to unity, it is difficult to discover whether these high estimates are due to a slowly growing component of the series or to long lags in adjustment.
2. Test for the possibility of misspecification of the model by including additional lagged terms of the exogenous variables.
3. If non-linear regression rountines are available, use them to test simultaneously for the presence of serial correlation.

[14/] James E. Martin, "The Use of Distributed Lag Models Containing Two Lag Parameters in the Estimation of Elasticities of Demand", Journal of Farm Economics, Vol. 45, No. 2, December 1963, pp. 1474-1481.

[15/] Zvi Griliches, "Distributed Lags: A Survey", Econometrica, Vol. 35, No. 1, January 1967, pp. 16-49.

4. Forget about the Durbin-Watson statistic in this context as a test for serial correlation in the original disturbances. It is very badly biased.
5. Do not expect the data to give a clear-cut answer about the exact form of the lag. One should try to get more implications from theory about the correct form of the lag and impose it on the data.
6. Interpret the coefficients of a distributed lag model with great care, since the same reduced form can arise from many different structures. Moreover different reduced forms may not differ much in the fit they provide to data, but have widely different implications as to the underlying structure that generated the data.
7. To get better answers to such complicated questions we need better data and much larger samples.

Polynomial Distributed Lags

A flexible approach for incorporating various powers of lag and various time periods of lag suggested by Almon[16/]. Her scheme is predicated on the Weirstrass theorem which states that a function continuous over a closed interval may be approximated by a polynomial. Originally Almon used Lagrangian interpolation techniques to estimate the polynomial. Cooper[17/] has developed a more direct and a slightly less complicated approach to estimating polynomial lags. Of course the complexity of estimating polynomial lags increases rapidly with the degree of the polynomial being considered

---

16/ Shirley Almon, "The Distributed Lag Between Capital Appropriations and Expenditures", Econometrica, Vol. 33, No. 1, January 1965, pp. 178-196.

17/ J. Phillip Cooper, "Two Approaches to Polynomial Distributed Lags Estimation: An Expository Note and Comment", The American Statistician, Vol. 26, No. 3, June 1972, pp. 32-35.

and the length of lag being considered. Both the length of lag and the degree of the polynomial distributed lag models and it is desireable to have both of these dimensions fairly small.

To illustrate the polynomial distributed lag model we shall draw on a supply model by Chen, Courtney, and Schmitz.[18/] Basically, the problem they were dealing with was one dealing with the response of milk production to changes in administered milk prices. Producer milk prices almost every where in the U. S. are regulated by either federal milk marketing orders or state milk control programs. Milk production can probably be decreases easier (sending dairy cows to the slaughter house) than it can be increased. To increase milk production certain biological constraints have to be considered. It is reasonable to expect changes in milk production to occur only after a certain amount of time has elapsed and also it's reasonable to consider the lagged response to be something other than linear. Thus, the polynomial distributed lag framework is a plausible framework for the type of problem being considered.

Let us consider the following supply response model.

$$(10.31) \quad q_t = \sum_{\tau=0}^{k} \beta_\tau P_{t-\tau} + \sum_{i=0}^{r} \gamma_i X_i + \varepsilon_t$$

where

q_t = output at time t.

$P_{t-\tau}$ = price at time $t-\tau$

k = number of periods covered by the lag function.

β_τ = the coefficients of the lag structure, i.e. β_0, β_1 - - - β_k. There are k + 1 of these coefficients to be determined.

γ_i = coefficients of the other variables beside price that determine output at time t.

18/ Dean Chen, Richard Courtney, and Andrew Schmitz, "A Polynomial Lag Formulation of Milk Response", American Journal of Agricultural Economics, Vol. 54, No. 1, February 1972, pp. 77-79.

X_i = variables other than price that determine output at time t.

Now to consider the polynomial distributed lag let us disregard the variables other thsn price and the disturbance term. Let us consider only:

$$(10.32) \quad q_t = \sum_{\tau=0}^{k} \beta_\tau \, P_{t-\tau}$$

In a generalized polynomial lag model the lagged weights are given by:

$$(10.33) \quad \beta_\tau = \alpha_0 + \alpha_1\tau + \alpha_2\tau^2 + - - - - + \alpha_N\tau^N$$

where N is the degree of the polynomial function. In this general form (10.32) could be rewritten as:

$$(10.34) \quad q_t = \sum_{\tau=0}^{k} (\alpha_0 + \alpha_1\tau + \alpha_2\tau^2 + - - - - + \alpha_N\tau^N) \, P_{t-\tau}$$

Generally a low ordered polynomial is considered adequate for most econometric work. A low order polynomial is considerably easier to handle computationally. For our illustration we shall consider a second order finite polynomial and the specific form of (10.34) is:

$$(10.35) \quad q_t = \sum_{\tau=0}^{k} (\alpha_0 + \alpha_1\tau + \alpha_2\tau^2) \, P_{t-\tau}$$

In this formulation β_τ is:

$$(10.36) \quad \beta_\tau = \alpha_0 + \alpha_1\tau + \alpha_2\tau^2$$

A further restriction on β_τ is that $\beta_\tau = 0$ when $\tau = k$. This yields

$$(10.37) \quad \alpha_0 + \alpha_1 k + \alpha_2 k^2 = 0$$

This restriction is imposed since it is assumed that beyond some length of time period k, price changes would no longer affect current production. Now solving (10.37) for α_0 yields:

$$(10.38) \quad \alpha_0 = -\alpha_1 k - \alpha_2 k^2$$

Substituting into (10.36) gives:

$$(10.39) \quad \beta_\tau = -\alpha_1 k - \alpha_2 k^2 + \alpha_1 \tau + \alpha_2 \tau^2$$

$$= \alpha_1(\tau - k) + \alpha_2 (\tau^2 - k^2)$$

Now (10.35) may be rewritten as:

$$(10.40) \quad q_t = \alpha_1 \sum_{\tau=0}^{k} (\tau - k) P_{t-\tau} + \alpha_2 \sum_{\tau=0}^{k} (\tau^2 - k^2) P_{t-\tau}$$

In (10.40) the price variables for which coefficients are estimated are:

$$(10.41) \quad \sum_{\tau=0}^{k} (\tau - k) P_{t-\tau}$$

and

$$(10.42) \quad \sum_{\tau=0}^{k} (\tau^2 - k^2) P_{t-\tau}$$

For estimation purposes (10.40) could be expanded to include the non-price variables in (10.31).

Note that (10.41) and (10.42) will tend to give negative weights because the number of time periods covered by the lag function is a positive number while τ varies from zero to k but never exceeds k. This makes no difference in terms of estimation. However, it may be more convenient to change the negatives of (10.41) and (10.42) in order to work with positive weights. This is accomplished by letting:

$$(10.43) \quad W_1 = \sum_{\tau=0}^{k} (k - \tau) P_{t-\tau}$$

and

$$(10.44) \quad W_2 = \sum_{\tau=0}^{k} (k^2 - \tau^2) P_{t-\tau}$$

Earlier we had assumed that a second order polynomial was the appropriate polynomial to consider. Now the number of periods covered by the lag is still an open question. However, prior to estimation the number

of periods, k, has to be selected. For illustrative purposes let us arbitrarily choose k=6. The weights are:

$$(10.45) \quad W_1 = 6P_t + 5P_{t-1} + 4P_{t-2} + 3P_{t-3} + 2P_{t-4} + P_{t-5}$$

and

$$(10.46) \quad W_2 = 36P_t + 35P_{t-1} + 32P_{t-2} + 27P_{t-3} + 20P_{t-4} + 11P_{t-5}$$

Now the equation to be estimated is (10.40) with (10.45) and (10.46) substituted into it. Thus the equation of interest is

$$(10.47) \quad q_t = \alpha_1 W_1 + \alpha_2 W_2$$

and hence,

$$(10.48) \quad q_t = \alpha_1 (6P_t + 5P_{t-1} + 4P_{t-2} + 3P_{t-3} + 2P_{t-4} + P_{t-5}) + \alpha_2 (36P_t + 35P_{t-1} + 32P_{t-2} + 27P_{t-3} + 2P_{t-4} + P_{t-5})$$

In estimation our interest is to obtain estimates of α and α_2, $\hat{\alpha}_1$ and $\hat{\alpha}_2$, respectively. To obtain estimates of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ ordinary least squares can be applied to (10.48).

If the appropriate model contains variables in addition to the price variable, then the model of interest is (10.31). Ordinary least squares may be applied to:

$$(10.49) \quad q_t = \alpha_1 W_1 + \alpha_2 W_2 + \gamma_1 X_1 + \gamma_2 X_2 + - - - + \gamma_r X_r + \epsilon_t$$

to obtain estimates of the linear and quadratic components of the lag and the parameters of the other predetermined variables. The values for W_1 and W_2 would of course be obtained from (10.45) and (10.46) respectively.

Once estimates of α_1 and α_2 are obtained the linear and quadratic effects of the price for each of periods, current to t-5, may be computed and the total effect (linear and quadratic) of each of the periods, current to t-5 may be evaluated. Also, one may have alternative hypotheses about

the plausible periods of lag, which may be assessed. Of course to do this a separate equation will have to be estimated for each alternative hypothesis about the lag.

Finally, note that β_τ was the parameter of the price variable in the original model but we ended up estimating α's. From (10.39) we see that:

$$(10.50) \quad \beta_\tau = \alpha_1 (\tau - k) + \alpha_2 (\tau^2 - k^2)$$

If we wish to test hypotheses about β_τ directly we need the standard error of β_τ. The variance of β_τ is:

$$(10.51) \quad \text{Var}(\beta_\tau) = \text{Var}\left[\alpha_1 (\tau - k) + \alpha_2 (\tau^2 - k^2)\right]$$

$$= (\tau - k)^2 \text{Var}(\alpha_1) + (\tau^2 - k^2)^2 \text{Var}(\alpha_2)$$

$$+ 2(\tau - k)(\tau^2 - k^2) \text{Cov}(\alpha_1 \alpha_2)$$

and

$$(10.52) \quad t = \frac{\beta_\tau}{\sqrt{\text{Var}(\beta_\tau)}}$$

with degrees of freedom associated with the residual variance of (10.48) or (10.49), whichever was used. About the same information may be obtained by testing the α's and γ's in (10.49).

Chapter 11. Models and Identification

Introduction

Thus far we have been concerned mainly with single equations and focus has been on estimation within the least squares framework and testing of hypotheses. We have not concerned ourselves with whether or not the equation we were estimating represented some type of structural relationship. In fact we really haven't worried very much about how and why the relationship of interest is what it is. As a point of departure we took the equation as given and concerned ourselves with estimation and testing of hypotheses about certain parameters or sets of parameters. We assumed that the equation represented the structure of interest or that the equation was a prediction equation and the underlying economic structure was of no interest. Furthermore, the types of relationships with which we were concerned consisted of one endogenous variable and a set of predetermined variables which were assumed to truly be predetermined. Under these conditions the issues we are going to be discussing didn't arise.

In this chapter attention will first be devoted to putting the model in perspective. Then focus will be on various ramifications of identification. Identification is one of the more important concepts to understand with respect to empirical economic analysis. In the early sections it will be helpful to define a number of concepts which are essential to understanding identification and simultaneous equations systems.

Structure and Economic Models

Much of econometric analysis is concerned with the estimation of the parameters of structural relationships. In other words estimating the parameters of an economic structure. Structure is the relationship or set of relationships capable of generating the numerical observations in which we are interested.

Model is somewhat of a broader concept. A model is a specification of classes of structures. An economic model is theoretical in nature although experience, previous studies, and observation may have been used in selecting the economic theory or theories to be used. The model will entail a combination of behavioral, technical, institutional and definitional relationships. Some of these relationships may be identities particularly definitional or institutional relationships. The model includes the number of equations, the general form of the equations, and a priori restrictions the parameters of the relationships.

The model is sort of an ideal representation of the relationship or relationships in which we are interested. The variables in the model are ideal. They are what economic theory tells us ought to be included if the model is to make sense. The variables are ideal in the sense that they are exactly what is needed to explain the phenomena of interest and they represent exactly what they are supposed to represent. The relationships in the economic model are usually expressed in general form and functional form issues are not of major concern.

Statistical Model and Variables

Before empirical analyses may be performed specific functional forms of the relationships in the economic model have to be specified and "real world" variables which represent the ideal economic variables need to be selected and used. The economic model has to be translated into a statistical model. Estimation and empirical analysis are performed on the statistical model which provide the basis for conclusions and inferences about the particular economic model from which the statistical model evolved.

The statistical model is basically a stochastic representation of the economic model. It contains a stochastic disturbance term. The stochastic disturbance term is simply an explicit recognition that the

relationships we have specified don't hold exactly all the time, if ever. There may be any one or more of several reasons for this. First even though the economic theory we are using may be "good" and exact, measurement problems may occur when we consider specific functional forms of relationships and specific empirical variables in the statistical model. The specific functional forms chosen in the statistical model may not be and very likely will not be exactly the true functional forms of the economic phenomena of interest. And we have no way of knowing what the exact functional forms of the economic phenomena are. Even if we did know, the true functional forms might be such complex curvilinear forms that we would need to use some simpler curvilinear or linear functional forms as approximations. The variables which are included in the statistical model to represent the economic variables in the economic model may not be measuring exactly the economic phenomena they are supposed to be and insofar as they don't measurement errors with regard to the variables arise. Second, in economic and statistical model specification we do not attempt to include in the model all variables which might affect the particular variable or variables of interest. Generally, those believed to be the major ones are explicitly included and hence some are "left out". The composite effects of the "left out" variables are assumed to be random and the stochastic disturbance term may be representing these "left out" variables. Third, some behavior of people and other economic units may in fact be random and if so the economic and statistical models would not hold exactly. The random disturbance term may be representing some combination of the above and if any or all actually occur then the empirical model needs to be specified with some type of random disturbance term.

Before the theoretical constructs can be confronted with empirical evidence either for estimation or testing purposes we have to commit

ourselves to specific functional forms of equations representing the relationships in the economic model. Generally, a simple functional form is preferred to a complex functional form. If the more complex functional form more closely approximates the functional form believed to be true than does the simpler form then one is faced with making a choice about the trade-off between the simpler form which may result in greater error and the more complex form which likely will result in less errors. The more complex functional form may raise some severe problems in estimation and analyses. One always tries to choose functional forms which are not inconsistent with economic behavior and the particular economic theories or theories drawn upon in developing the economic model. This is particularly relevant when the choice is among functional forms that are non-linear in variables.[1/] It may be possible to estimate the parameters of some non-linear functional forms, various types of exponential forms; however, the functional forms may exhibit various characteristics in certain ranges of the variables which are not possible if the relationship is supposed to represent some kind of economic behavior. Furthermore, some functional forms which appear to exhibit some desirable properties may be next to impossible and in some cases impossible to work with empirically in terms of obtaining estimates of the desired parameters. Again great care has to be exercised especially when dealing with functional forms which are non-linear in variables and functional forms which are non-linear in parameters. Finally, one always wants to choose a functional form which lends itself to the types of analyses required.

[1/] The same would apply for functional forms which are non-linear in parameters. In fact the problems are even more severe. Estimation of values of parameters of functional forms which are non-linear in parameters is complicated and difficult even in cases where it is possible. Insofar as possible we will steer away from models entailing functional forms that are non-linear in parameters.

Clearly a functional form which does not lend itself to the type of analyses desired is not very useful even if it is possible to obtain estimates of the parameters of the function form. For example, if marginal concepts form an integral part of the desired economic analysis then the choice of functional forms should be a choice among functional forms which can provide information about the required marginal quantities.

The variables specified in the theoretical economic model usually don't exist as such. Among the existing variables which more or less approximate the variable or variables in the economic model one has to choose the most appropriate serrogate. Frequently several alternatives are available for a particular variable. For example in one of the structures of the economic model the price of a given commodity may be specified as a key variable in that structure. There may be several price series of this commodity available or one or more price indices of the commodity none of which is measuring exactly the phenomena that was specified by the corresponding variable in the economic model. Often the differences stem from some aspect of aggregation. Nevertheless one has to choose empirical variables which most closely measure the ones specified in the theoretical economic model. More often than not this choice is made with less than adequate guides or criteria.

Classification of Variables

The variables in the economic and statistical model have to be classified into endogenous and predetermined variables. The endogenous variables are those whose values are determined within the system we are considering. Their values are determined by the model we have specified and its underlying economic theory. The values of the predetermined variables are either determined within the model but at some previous time period or outside of the model we have specified at either current or previous time periods. The predetermined variables consists

of two types of variables - lagged èndogenous and exogenous. The values of lagged endogenous variables are determined within the economic model but at a previous time period. In other words from the viewpoint of the current time period the values of the lagged endogenous variables are "history" and we take their values as given. The values of the exogenous variables are determined outside the model under consideration within the current time period or during some previous time period.

The necessity for classifying the variables in the model into endogenous and predetermined is that the number of the different kinds of variables we have in the model along with the algebraic forms of the equations will determine whether we can estimate the parameters of interest and the kind of estimation procedures which will have to be employed.

Role of the Model in Determining Estimation Method

The statistical model depends directly upon the economic model. The economic model gives some guides with respect to the form of the relationship and the variables to be included in the relationships. The economic model is the basis for determining causality and the flow of causal forces. The economic model provides information about the number of equations in the system being considered. The economic model provides information about the restriction on the coefficients of the relationships. These restrictions may or may not be directly incorporated in the statistical model and in the estimation of the parameters of the statistical model. It is from the economic model that we determine that changes in certain variables in the relationships of the model cause changes in certain other variables in the model.

The statistical model follows from the economic model. The statistical model specifies the relationships of the economic model in the specific algebraic forms and specific measures of the variables in the economic model are incorporated in the statistical model. The statistical

model provides no direct information about the flows of causal forces. The statistical model provides information about the associations of variables. Information about associations among variables in the statiatical model is supporting evidence about the flows of causal forces in the economic model. In other words conclusions about the associations among variables are made from the atatistical model. But these conclusions are interpreted through the economic model which enables the information about associations to be translated into causality. It is from the economic model that we draw our conclusions and inferences about causality; however, the information about associations from the statistical model is utilized.

The translation of atatistical resulta through the economic model to draw conclusions and inferences is an important consideration in the estimation of the parameters of the statistical model. We want to perform the estimation in a way that is amenable to such a translation. This means we want to know the exact relationships between our statistical model and our economic model. If the relationships in the economic model represent particular economic structures we want to estimate the parameters of the corresponding statistical equations in such a way that the estimators provides estimates of the parameters of interest or provide estimates of coefficients from which estimates of the parameters of interest may be derived.

The appropriate method of estimation of parameters of the statistical model depends on the number of equations in the model and the numbers of endogenous and predetermined variables in the equations. It is the combination of equations in the model and variables within equations which give rise to identification issues. The model may consist of 1) one equation; one endogenous variable, 2) several equations; only one endogenous variable in each equation; 3) one equation; more than one

endogenous variable, or 4) several equations; more than one endogenous variable in each equation. The type of identification will indicate whether or not it will be possible to estimate the parameters of interest. Second, if it is possible to obtain estimates of structural parameters the type of identification suggests the general type of estimation procedure which will need to be employed. Specifically it indicates whether single equation methods or simultaneous-systems-of-equations methods will need to be employed.

Completeness of a Model

In econometric analysis we are concerned with complete models. Complete when used in this manner has a very special meaning. A complete model is one in which there is one equation for each endogenous variable in the model. The equations may represent behavioral relationships (supply, demand, consumption relations, etc.), technical relationships (production functions, etc.), institutional relationships (tax rates, exchange rates, etc.) or definitional identities (market clearing conditions, per capita transformations, etc.).

The completeness is only with reference to the particular model specification being considered and only in the sense that there are as many equations in the model as there are endogenous variable. Insofar as the model which has been specified includes all structural relationships and variables which explain a particular economic phenomena the model is complete in terms of an explanation of the economic phenomena. But if the specified model, for one reason or another, does not include all of the relevant structural relationships and variables then the model may be complete in terms of containing as many equations as endogenous variables yet it may be incomplete from the viewpoint of explaining the economic phenomena in all of its possible dimensions.

Reduced Form Equations

Reduced form equations express one endogenous variable in a system of equations as a function of all predetermined variables (exogenous and lagged endogenous) in the system. In a model consisting of a system of equations representing structural relationships the reduced form are a particular type of transformation of the structural relationships. The structural relationships are transformed in such a way that there are as many resulting equations as there are endogenous variables in the system and in each equation one endogenous variable is a function of all predetermined variables in the entire system. Identities are part of the entire system.

The set of reduced form equations for a given set of structural equations is unique. Given a particular structural model there is only one set of reduced form equations for that structure. However, the converse is not true. Given a set of reduced form equations the structural model or models to which they pertain is not unique. More than one set of structural equations may have the same set of reduced form equations. This property of reduced form equations plays quite a crucial role in identification and in estimation of simultaneous systems of equations.

The least squares single equations with which we have concerned ourselves thus far are generally reduced form equations. They have been of the form one endogenous variables as a function of a set of predetermined variables. If the single equation by itself did not represent some type of economic structure then it represented some type of transformation of some part of an economic structure. Hence, it may have been one reduced form from a set of reduced form equations. But least squares equations are not always reduced form equations. When the least squares regression equations really represent an economic structure then even though they are single equations they are structural

equations. A structural model may consist of one or a set of single equations each containing only one endogenous variable.

Identification

Identification refers to whether or not the observations we have can provide us with information about the parameters of the structural equations of the model in which we are interested. Specifically it is concerned with whether or not we can go from the reduced form equations to the structural equations.

Identification is a property of the model - whether the model has the power of being structure identifying. It has to do with the amount of prior information in the model. Identification is a relative concept. The identification power of a model can vary and some models have more structure identifying power while others have less. The structure identifying power of a model is ultimately probabilistic in nature in that identification depends on certain parameters of a model being different from zero and these parameters are judged to be different from zero on a probability basis. Hence, a parameter which is judged different from zero at a lower probability level is less structure identifying than if it were judged different from zero at a higher probability. Hence, if a model is structure identifying, it is structure identifying in a relative sense.

Some models are not structure identifying at all. In other words the model is under-identified. This is a conceptual issue concerned with the prior information in the model and not an empirical issue. No amount of empirical data in terms of more observations or increased sample size will lead to identification of an under-identified model. The problem lies witht the model specification itself. The model just doesn't contain enough information to provide estimates of the structural parameters.

However, identification is viewed not only for a model as a whole but for each equation in the model and each parameter in each structural equation. A model as a whole may not be structure identifying but certain structural equations may be identified. Furthermore, certain structural equations may not be identified but some of the parameters of these equations may be identified. In other words the model may contain sufficient prior information to provide information about some parameters of some of the structural equations but it does not contain sufficient prior information to provide information about all parameters of all structural equations. The model as a whole is not structure identifying but some structural parameters are identified. From an estimation viewpoint this means that these particular parameters could be estimated given the particular model. If these happen to be the ones of interest then they could be estimated even though the model as a whole is not structure identifying.

Identification arises as a problem only when we are interested in structure. If interest is only in prediction of values of certain variables and the underlying economic structure is of no concern, then identification is not an issue. From a statistical point of view, identification arises only when the observations or data are governed by more than one equation. This is why biological and other physical scientists who used statistical tools didn't concern themselves with identification problems. The data with which they were working consisted of one endogenous variable and a set of predetermined variables and the structure could be represented by a single equation. Insofar that they were correct about the one endogenous variable and one equation structure they faced no identificarion problems. The problem of identification really came to the fore front when economist and other social scientist started applying statistical tools to data generated by an economy or a

sector of the economy where the data were believed to be governed by sets of multiple equations. It was apparent that many observationally equivalent relationships could be obtained from this type of data. The observations were generated by some joint probability distributions. From a statistical viewpoint the observationally equivalent relationships are indistinguishable and really make no difference. But from an economic viewpoint they are not equivalent. A supply equation is not the same as a demand equation and neither is the same as some "mongrel" or "bogus" equation which is some linear combination of equations.

Simple Supply and Demand Model

To explore the concept of identification, let us consider the following simple supply and demand model:

$$\begin{aligned} &\text{D:}\quad Y_{t1} = \beta_0 + \beta_1 Y_{t2} + u_{t1} \\ &\text{S:}\quad Y_{t1} = \alpha_0 + \alpha_1 Y_{t2} + u_{t2} \end{aligned} \tag{11.1}$$

where: Y_{t1} is quantity

Y_{t2} is price

u_{t1} and u_{t2} are random disturbances

β_0, β_1, α_0, and α_1 are parameters.

Let us assume that $\beta_1 < 0$ and $\alpha_1 > 0$. These maintained hypotheses contain little information which will help in estimating the structural parameters. The more a priori information or restrictions which can be incorporated into the maintained hypotheses (model), that is more restrictions put on the data, the more specific the conclusions which can be made. However, of course, if incorrect restrictions are imposed then specific incorrect conclusions will be made.

Now let us approach the identification problem with reference to (11.1) from an observation viewpoint. Suppose we have only one observation of quantity and price, Y_1 and Y_2. The situation could be one of those presented in Figure 11.1, Figure 11.2, or Figure 11.3. In Figure 11.1, $u_1 = 0$. In Figure 11.2, $u_2 = 0$. In Figure 11.3, $u_1 = u_2 = 0$.

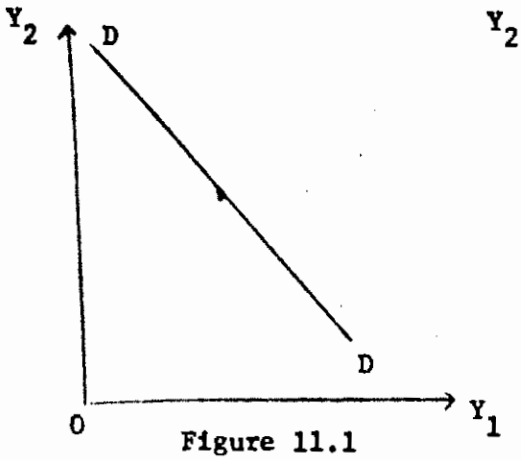


Figure 11.1

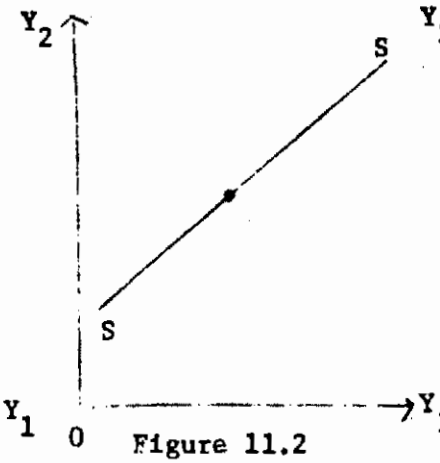


Figure 11.2

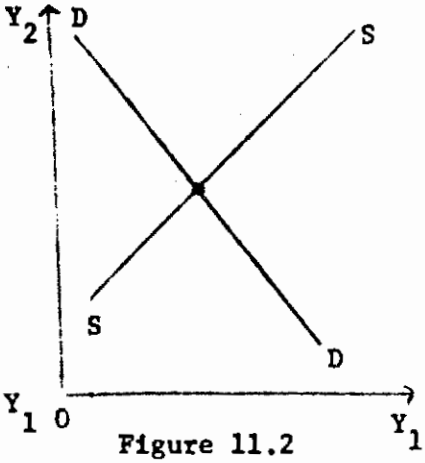


Figure 11.2

Given $u_1 = 0$, we know the observation is a point on the demand curve. Given $u_2 = 0$ we know the observation is a point on the supply curve. And given that $u_1 = u_2 = 0$ we know the observation is a point on both the demand and supply curves. But note that we have additional information about the disturbances which is the information that enabled us to say on which function the observation is a point. Obviously with only one observation we are in no position to obtain any kind of estimates of the structural parameters in (11.1) but the additional information about the disturbances enabled us to make more specific conclusions than we would have been able to make without it. The slopes of the curves in Figure 11.1, Figure 11.2, and Figure 11.3 were arbitrarily chosen and to obtain estimates of the slopes we need more observations.

Suppose that we now have two observations on quantity and prices. We could have any one of the situations presented in Figures 11.4-11.7. In Figure 11.4, $u_1 = 0$, in Figure 11.5, $u_2 = 0$ and in Figure 11.6, $u_1 = u_2 = 0$.

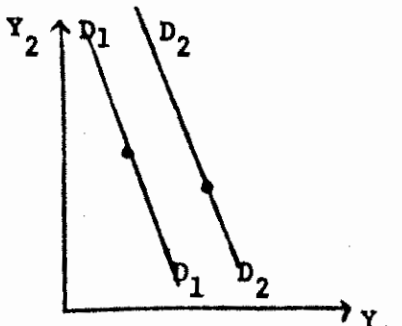


Figure 11.4

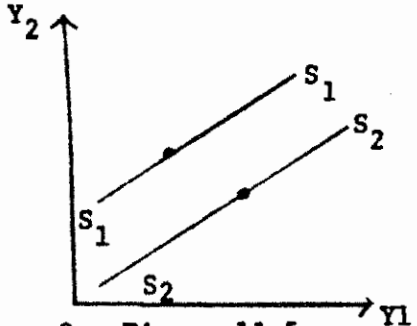


Figure 11.5

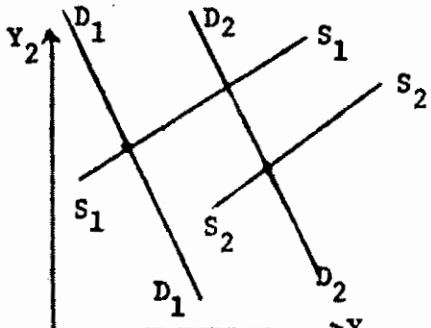


Figure 11.6

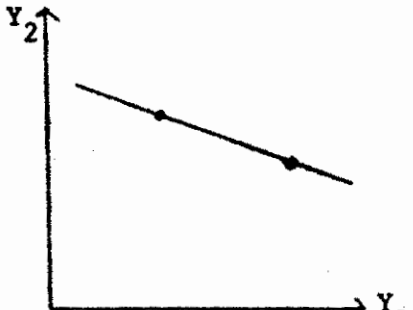


Figure 11.7

Note with the additional information about the disturbances we are in a position to say something about which functions the observations are a point on. But from a statistical estimation viewpoint we do not have adequate information to estimate the structural parameters of (11.1). There are four parameters and we have only two observations. However, with the information about the u's we were able to conceptually identify which curves the points pertained to.

If we were to utilize the two observations to fit a curve (straight line in this case) we would obtain the one given in Figure 11.7. Now if the supply and demand structures are as given in either of Figures 11.4-11.6 then the representation in Figure 11.7 is a "mongrel". It is the curve of a "bogus" equation. It is a linear combination of the structural equations. It is observationally equivalent to the structural equations in (11.1) but a mixture of both. From a statistical viewpoint, it is indistinguishable from either of the equations in (11.1) but certainly is not equivalent in an economic sense. Although we have been considering

only two observations, the caae illustrates what difficulties arise in estimating structural relationships when many observations are available. The problems arose because (11.1) is not a structure identifying model. We basically need more information in (11.1) about factors which affect demand separate from those which affect supply. Then we would be able to distinguish the structural equations from each other and if we had adequate observations we could obtain estimates of the structural parameters.

Let us further explore the characteristics of the "bogus" equation which is graphically presented in Figure 11.7. Let us take an arbitrary linear combination of the structural equations in (11.1). We have:

$$k_1 Y_{t1} = k_1 \beta_0 + k_1 \beta_1 Y_{t2} + k_1 u_{t1}$$

$$k_2 Y_{t1} = k_2 \alpha_0 + k_2 \alpha_1 Y_{t2} + k_2 u_{t2} \tag{11.2}$$

where k_1 and k_2 are some arbitrarily chosen constants not equal to zero. If we sum the two equations in (11.2) we have:

$$(k_1 + k_2) Y_{t1} = (k_1 \beta_0 + k_2 \alpha_0) + (k_1 \beta_1 + k_2 \alpha_1) Y_{t2} + (k_1 u_{t1} + k_2 u_{t2}) \tag{11.3}$$

and

$$Y_{t1} = \left(\frac{k_1 \beta_0 + k_2 \alpha_0}{k_1 + k_2} \right) + \left(\frac{k_1 \beta_1 + k_2 \alpha_1}{k_1 + k_2} \right) Y_{t2} + \left(\frac{k_1 u_{t1} + k_2 u_{t2}}{k_1 + k_2} \right) \tag{11.4}$$

and

$$Y_{t1} = \gamma_0 + \gamma_1 Y_{t2} + w_t \tag{11.5}$$

where

$$\gamma_0 = \frac{k_1 \beta_0 + k_2 \alpha_0}{k_1 + k_2}$$

$$\gamma_1 = \frac{k_1 \beta_1 + k_2 \alpha_1}{k_1 + k_2}$$

$$w_t = \frac{k_1 u_{t1} - k_2 u_{t2}}{k_1 + k_2}$$

Note how similar (11.5) is to the equations in (11.2) and (11.1). If we had numerical values of the coefficients in (11.5) we couldn't distinguish whether they were really coefficients of (11.5) or coefficients of either of the equations in (11.2) or coefficients of either equations in (11.1).

A Model With More Information In The Demand

Let us now explore the identification of a supply and demand model which has one predetermined variable in the demand relationship. The model to be considered is:

(11.6) $$D: \quad Y_{t1} = \beta_0 + \beta_1 Y_{t2} + \beta_2 Z_{t1} + u_{t1}$$

$$S: \quad Y_{t1} = \alpha_0 + \alpha_1 Y_{t2} + u_{t2}$$

$$Y^D_{t1} = Y^S_{t1}$$

where Y_{t1} is the quantity of the product in t.

Y_{t2} is the price of the product in t.

Z_{t1} is some measure of income in t.

u_{t1} and u_{t2} are disturbances.

α's and β's are structural parameters.

Z_{t1} is a predetermined variable and it is assumed that $\beta_1 < 0$, $\beta_2 > 0$, and $\alpha_1 > 0$. $(Y^D_{t1} = Y^S_{t1})$ is an identity denoting a market clearing condition which indicates the observed quantities pertain to equilibrium situations where quantity demanded is equal to quantity supplied.

Now let us explore the identification properties of (11.6). Taking a linear combination of (11.6) we have:

(11.7) $$k_1 Y_{t1} = k_1 \beta_0 + k_1 \beta_1 Y_{t2} + k_1 \beta_2 Z_{t1} + k_1 u_{t1}$$

$$k_2 Y_{t1} = k_2 \alpha_0 + k_2 \alpha_1 Y_{t2} + k_2 u_{t2}$$

$$Y^0_{t1} = Y^S_{t1}$$

and

(11.8) $$(k_1 + k_2) Y_{t1} = (k_1 \beta_0 + k_2 \alpha_0) + (k_1 \beta_1 + k_2 \alpha_1) Y_{t2} + k_1 \beta_2 Z_{t1} + (k_1 u_{t1} + k_2 u_{t2})$$

and

$$(11.9)\quad Y_{t1} = \left(\frac{k_1 \beta_0 + k_2 \alpha_0}{k_1 + k_2}\right) + \left(\frac{k_1 \beta_1 + k_2 \alpha_1}{k_1 + k_2}\right) Y_{t2} + \frac{k_1 \beta_2}{k_1 + k_2} Z_{t2} + \left(\frac{k_1 u_{t1} + k_2 u_{t2}}{k_1 + k_2}\right)$$

Hence:

$$(11.10)\quad Y_{t1} = \gamma_0 + \gamma_1 Y_{t2} + \gamma_2 Z_{t1} + w_t$$

where

$$\gamma_0 = \frac{k_1 \beta_0 + k_2 \alpha_0}{k_1 + k_2}$$

$$\gamma_1 = \frac{k_1 \beta_1 + k_2 \alpha_1}{k_1 + k_2}$$

$$\gamma_2 = \frac{k_1 \beta_2}{k_1 + k_2}$$

$$w_t = \frac{k_1 u_{t1} + k_2 u_{t2}}{k_1 + k_2}$$

The equation (11.10) is a "bogus" equation. It is observationally equivalent to the demand equation in (11.6) but note it is not observationally equivalent to the supply equation. Thus by having more information in the demand equation in terms of a predetermined variables, assuming $\beta_2 \neq 0$, we have achieved identification of the supply relation.

Now let us obtain the reduced form equations for the structures in (11.6) We want to express Y_{t1} and Y_{t2} as functions of all predetermined variables in the system. In this case Z_{t1} is the only predetermined variable. Let us solve for Y_{t2} from the supply equation and we have:

$$(11.11)\quad Y_{t2} = \frac{Y_{t1}}{\alpha_1} - \frac{\alpha_0}{\alpha 1} - \frac{u_{t2}}{\alpha_1}$$

Substituting into the demand equation is (11.6) we have

$$(11.12)\quad Y_{t1} = \beta_0 + \beta_1 \left(\frac{Y_{t1}}{\alpha_1} - \frac{\alpha_0}{\alpha_1} - \frac{u_{t2}}{\alpha_1} \right) + \beta_2 Z_{t1} + u_{t1}$$

$$= \left(\beta_0 - \frac{\beta_1 \alpha_0}{\alpha_1} \right) + \frac{\beta_1 Y_{t1}}{\alpha_1} + \beta_2 Z_{t1} + \left(u_{t1} - \frac{\beta_1 u_{t2}}{\alpha_1} \right)$$

and

$$(11.13)\quad Y_{t1} - \frac{\beta_1 Y_{t1}}{\alpha_1} = \left(\frac{\alpha_1 \beta_0 - \beta_1 \alpha_0}{\alpha_1} \right) + \beta_2 Z_{t1} + \left(\frac{\alpha_1 u_{t1} - \beta_1 u_{t2}}{\alpha_1} \right)$$

$$\left(\frac{\alpha_1 - \beta_1}{\alpha_1} \right) Y_{t1} = \left(\frac{\alpha_1 \beta_0 - \beta_1 \alpha_0}{\alpha_1} \right) + \beta_2 Z_{t1} \left(\frac{\alpha_1 u_{t1} - \beta_1 u_{t2}}{\alpha_1} \right)$$

hence

$$(11.14)\quad Y_{t1} = \left(\frac{\alpha_1 \beta_0 - \beta_1 \alpha_0}{\alpha_1 - \beta_1} \right) + \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1} Z_{t1} + \left(\frac{\alpha_1 u_{t1} - \beta_1 u_{t2}}{\alpha_1 - \beta_1} \right)$$

or

$$(11.15)\quad Y_{t1} = \pi_{10} + \pi_{11} Z_{t1} + v_{t1}$$

where

$$\pi_{10} = \frac{\alpha_1 \beta_0 - \beta_1 \alpha_0}{\alpha_1 - \beta_1}$$

$$\pi_{11} = \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1}$$

$$v_{t1} = \frac{\alpha_1 u_{t1} - \beta_1 u_{t2}}{\alpha_1 - \beta_1}$$

(11.14) or (11.15) is the reduced form equation for Y_{t1}. Y_{t1} is a function of Z_{t1} and Z_{t1} is the only predetermined variable in the system.

In a similar manner, let us solve for Y_{t2}. We could use either the supply or demand equation to solve for Y_{t2}, then substitute for Y_{t1} from the other equation. Let us use the supply equation:

$$(11.16)\quad Y_{t2} = \frac{Y_{t1}}{\alpha_1} - \frac{\alpha_0}{\alpha_1} - \frac{u_{t2}}{\alpha_1}$$

Substituting for Y_{t1} from the demand equation in (11.6) we have

$$(11.17) \quad Y_{t2} = \frac{\left(\beta_0 + \beta_1 Y_{t2} + \beta_2 Z_{t1} + u_{t1}\right)}{\alpha_1} - \frac{\alpha_0}{\alpha_1} - \frac{u_{t2}}{\alpha_1}$$

$$= \frac{\beta_0}{\alpha_1} + \frac{\beta_1}{\alpha_1} Y_{t2} + \frac{\beta_2}{\alpha_1} Z_{t1} + \frac{u_{t1}}{\alpha_1} - \frac{\alpha_0}{\alpha_1} - \frac{u_{t2}}{\alpha_1}$$

$$= \left(\frac{\beta_0 - \alpha_0}{\alpha_1}\right) + \frac{\beta_1}{\alpha_1} Y_{t2} + \frac{\beta_2}{\alpha_1} Z_{t1} + \left(\frac{u_{t1} - u_{t2}}{\alpha_1}\right)$$

and,

$$(11.18) \quad \left(\frac{\alpha_1 - \beta_1}{\alpha_1}\right) Y_{t2} = \frac{\beta_0 - \alpha_0}{\alpha_1} + \frac{\beta_2}{\alpha_1} Z_{t1} + \frac{u_{t1} - u_{t2}}{\alpha_1}$$

hence,

$$(11.19) \quad Y_{t2} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} Z_{t1} + \frac{u_{t1} - u_{t2}}{\alpha_1 - \beta_1}$$

or

$$(11.20) \quad Y_{t2} = \pi_{20} + \pi_{21} Z_{t1} + v_{t2}$$

where

$$\pi_{20} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$$

$$\pi_{21} = \frac{\beta_2}{\alpha_1 - \beta_1}$$

$$v_{t2} = \frac{u_{t1} - u_{t2}}{\alpha_1 - \beta_1}$$

Now from the reduced form equations for (11.6) we have four reduced form parameters each of which is a function of the structural parameters.

$$(11.21)\quad \pi_{10} = \frac{\alpha_1\beta_0 + \beta_1\alpha_0}{\alpha_1 - \beta_1} \qquad \pi_{11} = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}$$

$$\pi_{20} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \qquad \pi_{21} = \frac{\beta_2}{\alpha_1 - \beta_1}$$

But there are five structural parameters which are unknowns - β_0, β_1, β_2, α_0, and α_1. We do not have adequate information to determine values of all five structural parameters even though we may be able to estimate values of all of the reduced parameters. Even though the model as a whole is not structure identifying an equation or certain parameters of one or more of the equations may be identified thus enabling estimates of the identified parameters to be obtained.

In the particular case being considered the difficulty stems from the demand equation of the model. The part of the maintained hypothesis pertaining to the demand structure contains enough a priori information to make the supply equation identified but the supply equation doesn't contain enough a priori information to allow identification of the demand equation. Recognize that the Z_{t1} variable, income, in the demand equation which is permitting the supply equation to be identified. This is contingent on the fact that $\beta_2 \neq 0$.

Given that we have achieved some type of identification of the supply equation let us attempt to derive the parameters of the supply equation from the reduced form parameters presented in (11.21). We can see that α_1 can easily be obtained from π_{11} and π_{21} as follows

$$(11.22)\quad \alpha_1 = \frac{\pi_{11}}{\pi_{21}} = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1} \cdot \frac{\alpha_1 - \beta_1}{\beta_2}$$

Note that the reduced form parameters of (11.15) and (11.20) may be estimated by least squares methods. Hence,

$$(11.23) \quad \hat{\alpha}_1 = \frac{\hat{\pi}_{11}}{\hat{\pi}_{21}}$$

where $\hat{\pi}_{11}$ and $\hat{\pi}_{21}$ are least squares estimators and their respective estimates may be obtained by regressing Y_{t1} and Y_{t2} on Z_{t1} respectively.

By this method, however, we cannot derive α_0, β_0, β_1, and β_2 and not obtain estimate of them from estimated values of the reduced form parameters. Note that π_{10} and π_{20} each contain β_0, β_1, α_0, and α_1 and even though we can solve for α_1 this still leaves us with an indeterminate situation with respect to α_0 and the other structural parameters, other than α_1. α_1 is the only structural parameter which can be uniquely determined from the reduced form parameters. So when we indicated that we have identification of the supply equation we are really saying that the coefficient of the price variable, Y_{t2}, in the supply equation may be obtained uniquely from the reduced form parameters.

It may seem contradictory that the supply equation is identified yet we are unable to obtain α_0 from the reduced form parameters. However, this is really not serious. The important parameters are the coefficients of the variables not the intercepts which indicate the positions of the surfaces in space. Often times econometric models will be formulated with variables measured as deviations from their respective means and intereepts don't enter the model. But in this case if we are going to obtain an estimate of the intercept of the supply equation we are going to have to rely on additional information and estimate the intercept accordingly. One can position the supply equation so as to pass through the centroid of the data. This is consistent with a "best fitting" curve or surface where criterion of goodness of fit is a minimum sum of squared residuals which in this case would be a minimum $\sum_{t=1}^{n} \hat{\mu}_{t1}^2$. Utilizing this type of prior information about "best fit" and incorporating it as a constraint we obtain an estimate of α_0 in the following manner:

(11.24) $\hat{\alpha}_0 = \bar{Y}_{t1} - \hat{\alpha}_1 \bar{Y}_{t2}$

where $\bar{Y}_{t1}$ is the sample mean of the quantity observations.

$\bar{Y}_{t2}$ is the sample mean of the price observations.

$\hat{\alpha}_1$ is the estimate of α_1 given in (11.23)

A Model With More Information in the Supply

Let us briefly consider the following model:

(11.25) D: $Y_{t1} = \beta_0 + \beta_1 Y_{t2} + u_{t1}$

S: $Y_{t1} = \alpha_0 + \alpha_1 Y_{t2} + \alpha_2 Z_{t2} + u_{t2}$

$Y^D_{t1} = Y^S_{t1}$

where: Y_{t1} is the quantity of the product in t.

Y_{t2} is the price of the product in t.

Z_{t2} is the price in t of a factor used in the production of Y_{t1}.

u_{t1} and u_{t2} are disturbances.

α's and β's are structural parameters.

Let us consider Z_{t2} to be a predetermined variable. Also, let us assume that $\beta_1 < 0$, $\alpha_1 > 0$, and $\alpha_2 < 0$. $(Y^D_{t1} = Y^S_{t1})$ is a market clearing condition.

In (11.25) we have additional information in the supply equation in terms of the predetermined variable Z_{t2} whereas in (11.6) the additional information was contained in the demand equation in terms of Z_{t1}. Without going through the details we did for (11.6), the demand equation in (11.25) is identified while the supply equation is not. The situation is just reversed from (11.6). β_1 may be derived from the reduced form parameters of the reduced form equations pertaining to (11.25) Estimates of the reduced form parameters may be obtained by least squares methods, i.e., regressing

Y_{t1} and Y_{t2} each on Z_{t2}. The other structural parameters of (11.25) cannot be obtained directly from the reduced form parameters because again there are four reduced form parameters and five unknown structural parameters. An estimate of β_0 may be obtained in the same manner as was the estimate of α_0 of (11.6) namely constraining the demand equation to pass through the centroid of the observations.

In summary additional information in the maintained hypotheses permitted some structure identification. In the cases considered additional a priori information in the demand equation in the form of a predetermined variable allowed the supply equation of the model to be identified. The inclusion of a predetermined variable in the supply equation resulted in structure identification of the demand equation.

Just Identified Model

Now let us consider a model which has just enough a priori information in the maintained hypotheses so that all structural equations and thus all structural parameters are identified. Such a model may be formulated by utilizing the demand equation of (11.6) and the supply equation of (11.25). Hence, let the model be:

(11.26) D: $Y_{t1} = \beta_0 + \beta_1 Y_{t2} + \beta_2 Z_{t1} + u_{t1}$

S: $Y_{t1} = \alpha_0 + \alpha_1 Y_{t2} + \alpha_2 Z_{t2} + u_{t2}$

$$Y^D_{t1} = Y^S_{t1}$$

where Y_{t1} is the quantity of the product t.

Y_{t2} is the price of the product in t.

Z_{t1} is some measure of income in t.

Z_{t2} is the price in t of a factor used in the production of Y_{t1}.

u_{t1} and u_{t2} are disturbances.

α's and β's are structural parameters.

$(Y^D_{t1} = Y^S_{t1})$ is a market clearing condition.

Z_{t1} and Z_{t2} will be considered predetermined variables. Theoretically, we would expect $\beta_1 < 0$, $\beta_2 > 0$, $\alpha_1 > 0$, and $\alpha_2 < 0$.

Now multiplying the structural equations of (11.26) by arbitrarily chosen constants k_1 and k_2 we have:

$$(11.27) \quad k_1 Y_{t1} = k_1 \beta_0 + k_1 \beta_1 Y_{t2} + k_1 \beta_2 Z_{t1} + k_1 u_{t1}$$

$$k_2 Y_{t1} = k_2 \alpha_0 + k_2 \alpha_2 Y_{t2} + k_2 \alpha_2 Z_{t2} + k_2 u_{t2}$$

Summing the two equations of (11.27) yields:

$$(11.28) \quad (k_1 + k_2) Y_{t1} = (k_1 \beta_0 + k_2 \alpha_0) + (k_1 \beta_1 + k_2 \alpha_1) Y_{t2} + k_1 \beta_2 Z_{t1} + k_2 \alpha_2 Z_{t2} + (k_1 u_{t1} + k_2 u_{t2})$$

and

$$(11.29) \quad Y_{t1} = \left(\frac{k_1 \beta_0 + k_2 \alpha_0}{k_1 + k_2}\right) + \left(\frac{k_1 \beta_1 + k_2 \alpha_1}{k_1 + k_2}\right) Y_{t2} + \left(\frac{k_1 \beta_2}{k_1 + k_2}\right) Z_{t1} + \left(\frac{k_2 \alpha_2}{k_1 + k_2}\right) Z_{t2} + \left(\frac{k_1 u_{t1} + k_2 u_{t2}}{k_1 + k_2}\right)$$

Note that (11.29) is not observationally equivalent to either the supply or demand equations of (11.26) and there is no reason why it should be confused with either of the equations (11.29) contains both Z_{t1} and Z_{t2} whereas the demand equation in (11.26) does not contain Z_{t2} and the supply equation does not contain Z_{t1}. Also, note that there is a predetermined variable in the supply equation of (11.26) which is not in the demand equation and vice-verse.

Solving for the reduced form equations

$$(11.30) \quad Y_{t2} = \frac{Y_{t1}}{\alpha_1} - \frac{\alpha_0}{\alpha_1} - \frac{\alpha_2}{\alpha_1} Z_{t2} - \frac{u_{t2}}{\alpha_1}$$

Substituting into the demand equation of (11.26) yields:

$$(11.31)\quad Y_{t1} = \beta_0 + \beta_1 \left(\frac{Y_{t1}}{\alpha_1} - \frac{\alpha_0}{\alpha_1} - \frac{\alpha_2}{\alpha_1} Z_{t2} - \frac{u_{t2}}{\alpha_1} \right) + \beta_2 Z_{t1} + u_{t1}$$

and

$$(11.32)\quad \left(\frac{\alpha_1 - \beta_1}{\alpha_1} \right) Y_{t1} = \left(\frac{\alpha_1\beta_0 - \beta_1\alpha_0}{\alpha_1} \right) - \frac{\beta_1 \alpha_2}{\alpha_1} Z_{t2} + \beta_2 Z_{t1} + \left(\frac{\alpha_1 u_{t1} - \beta_1 u_{t2}}{\alpha_1} \right)$$

thus

$$(11.33)\quad Y_{t1} = \left(\frac{\alpha_1 \beta_0 - \beta_1 \alpha_0}{\alpha_1 - \beta_1} \right) + \left(\frac{\beta_2 \alpha_1}{\alpha_1 - \beta_1} \right) Z_{t1} - \left(\frac{\beta_1 \alpha_2}{\alpha_1 - \beta_1} \right) Z_{t2} + \left(\frac{\alpha_1 u_{t1} - \beta_1 u_{t2}}{\alpha_1 - \beta_1} \right)$$

or

$$(11.34)\quad Y_{t1} = \pi_{10} + \pi_{11} Z_{t1} + \pi_{12} Z_{t2} + v_{t1}$$

where

$$\pi_{10} = \frac{\alpha_1 \beta_0 - \beta_1 \alpha_0}{\alpha_1 - \beta_1}$$

$$\pi_{11} = \frac{\beta_2 \alpha_1}{\alpha_1 - \beta_1}$$

$$\pi_{12} = \frac{\beta_1 \alpha_2}{\alpha_1 - \beta_1}$$

$$v_{t1} = \frac{\alpha_1 u_{t1} - \beta_1 u_{t2}}{\alpha_1 - \beta_1}$$

The π's are reduced form parameters.

Now we may solve for the reduced form equation for price by substituting the demand equation of (11.26) for Y_{t1} in (11.30) which yields:

$$(11.35)\quad Y_{t2} = \frac{\beta_1 + \beta_1 Y_{t2} + \beta_2 Z_{t1} + u_{t1}}{\alpha_1} - \frac{\alpha_0}{\alpha_1} - \frac{\alpha_2}{\alpha_1} Z_{t2} - \frac{u_{t2}}{\alpha_1}$$

and

$$(11.36)\quad \left(\frac{\beta_1 - \alpha_1}{\alpha_1}\right) Y_{t2} = \left(\frac{\beta_0 - \alpha_0}{\alpha_1}\right) + \frac{\beta_2}{\alpha_1} Z_{t1} - \frac{\alpha_2}{\alpha_1} Z_{t2} + \left(\frac{u_{t1} - u_{t2}}{\alpha_1}\right)$$

thus

$$(11.37)\quad Y_{t2} = \left(\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\right) + \left(\frac{\beta_2}{\alpha_1 \beta_1}\right) Z_{t1} - \left(\frac{\alpha_2}{\alpha_1 - \beta_1}\right) Z_{t2} + \left(\frac{u_{t1} - u_{t2}}{\alpha_1 - \beta_1}\right)$$

or

$$(11.38)\quad Y_{t2} = \pi_{20} + \pi_{21} Z_{t1} + \pi_{22} Z_{t2} + v_{t2}$$

where

$$\pi_{20} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$$

$$\pi_{21} = \frac{\beta_2}{\alpha_1 - \beta_1}$$

$$\pi_{22} = -\frac{\alpha_2}{\alpha_1 - \beta_1}$$

$$v_{t2} = \frac{u_{t1} - u_{t2}}{\alpha_1 - \beta_1}$$

Now the reduced form parameters expressed as functions of the structural parameters are:

$$(11.39)\quad \pi_{10} = \frac{\alpha_1\beta_0 - \beta_1\alpha_0}{\alpha_1 - \beta_1} \qquad \pi_{20} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$$

$$\pi_{11} = \frac{\beta_2 \alpha_1}{\alpha_1 - \beta_1} \qquad \pi_{21} = \frac{\beta_2}{\alpha_1 - \beta_1}$$

$$\pi_{12} = \frac{\beta_1 \alpha_2}{\alpha_1 - \beta_1} \qquad \pi_{22} = \frac{\alpha_2}{\alpha_1 - \beta_1}$$

We have six equations or expressions of π's and six unkown structural parameters - α_0, α_1, α_2, β_0, β_1, and β_2. Thus the structural parameters derived from the reduced form parameters are:

$$\alpha_1 = \frac{\pi_{11}}{\pi_{21}} \qquad \beta_1 = \frac{\pi_{12}}{\pi_{22}}$$

(11.40)

$$\alpha_2 = \pi_{12} - \frac{\pi_{11}\,\pi_{22}}{\pi_{21}} \qquad \beta_2 = \pi_{11} - \frac{\pi_{12}\,\pi_{21}}{\pi_{22}}$$

$$\alpha_0 = \pi_{10} - \frac{\pi_{20}\,\pi_{11}}{\pi_{21}} \qquad \beta_0 = \pi_{10} - \frac{\pi_{12}\,\pi_{20}}{\pi_{22}}$$

Note that the parameters of (11.34) and (11.38) may be estimated by classical least squares methods. Once estimates of the reduced form parameters are obtained the estimates of the structural parameters may be derived from them. This is the indirect least squares method of estimating the structural parameters of a just identified system of simultaneous equations.

Over-Identified Model

Once again let us consider a two equation model but one in which there are two predetermined variables in each equation and the two predetermined variables are different for each equation. We shall determine the identification properties of such a model, investigate the reduced form equations and parameters, and explore what guides might be suggested for estimation of the parameters of the structural equations.

Consider the following model

(11.41)

$$D: \quad Y_{t1} = \beta_0 + \beta_1 Y_{t2} + \beta_2 Z_{t1} + \beta_3 Z_{t3} + u_{t1}$$

$$S: \quad Y_{t1} = \alpha_0 + \alpha_1 Y_{t2} + \alpha_2 Z_{t2} + \alpha_3 Z_{t4} + u_{t2}$$

$$Y^D_{t1} = Y^S_{t1}$$

where Y_{t1} is the quantity of the product in t.

Y_{t2} is the price of the product in t.

Z_{t1} is some measure of income in t.

Z_{t2} is the price in t of a factor used in the production of Y_{t1}.

Z_{t3} is some measure of preferences in t.

Z_{t4} is the price in t of a second factor used in the production of Y_{t1}.

u_{t1} and u_{t2} are disturbances

α's and β's are structural parameters

$(Y^D_{t1} = Y^S_{t1})$ is a market clearing condition.

Z_{t1}, Z_{t2}, Z_{t3}, and Z_{t4} will be considered predetermined variables and we are assuming that we have some measures of these variables and observations are available. Let us assume that $\beta_1 < 0$, $\beta_2 > 0$, $\beta_3 > 0$, $\alpha_1 > 0$, $\alpha_2 < 0$, and $\alpha_3 < 0$.

Now multiplying the structural equations of (11.41) by arbitrarily chosen constants k_1 and k_2 we have:

$$(11.42) \quad k_1 Y_{t1} = k_1 \beta_0 + k_1 \beta_1 Y_{t2} + k_1 \beta_2 Z_{t1} + k_1 \beta_3 Z_{t3} + k_1 u_{t1}$$

$$k_2 Y_{t1} = k_2 \alpha_0 + k_2 \alpha_1 Y_{t2} + k_2 \alpha_2 Z_{t2} + k_2 \alpha_3 Z_{t4} + k_2 u_{t2}$$

Summing these two equations yields:

$$(11.43) \quad (k_1+k_2) Y_{t1} = (k_1\beta_0 + k_2\beta_0) + (k_1\beta_1+k_2\beta_1) Y_{t2} + k_1\beta_2 Z_{t1} + k_2\alpha_2 Z_{t2} + k_1\beta_3 Z_{t3} + k_2\alpha_3 Z_{t4} + (k_1 u_{t1} + k_2 u_{t2})$$

and

$$(11.44) \quad Y_{t1} = \left(\frac{k_1\beta_0 + k_2\alpha_0}{k_1 + k_2}\right) + \left(\frac{k_2\beta_1 + k_2\alpha_1}{k_1 + k_2}\right) Y_{t2} + \left(\frac{k_1\beta_2}{k_1 + k_2}\right) Z_{t1} + \left(\frac{k_2\alpha_2}{k_1 + k_2}\right) Z_{t2} + \left(\frac{k_1\beta_3}{k_1 + k_2}\right) Z_{t3} + \left(\frac{k_2\alpha_3}{k_1 + k_2}\right) Z_{t4} + \left(\frac{k_1 u_{t1} + k_2 u_{t2}}{k_1 + k_2}\right)$$

All four predetermined variables are in (11.44) whereas the supply and demand equations of (11.41) each contain only two predetermined variables. Hence, (11.44) is not observationally equivalent to either of the two equations in (11.41). Also, note that there are two predetermined variables in the demand

equation of (11.41) which do not appear in the supply equation and vice-versa.

Solving for the reduced form equations:

$$(11.45) \quad Y_{t2} = \frac{Y_{t1}}{\alpha_1} - \frac{\alpha_0}{\alpha_1} - \frac{\alpha_2}{\alpha_1} Z_{t2} - \frac{\alpha_3}{\alpha_1} Z_{t3} - \frac{u_{t2}}{\alpha_1}$$

and substituting into the demand equation of (11.41) yields:

$$(11.46) \quad Y_{t1} = \left(\frac{\alpha_1 \beta_0 - \beta_1 \alpha_0}{\alpha_1 - \beta_1}\right) + \left(\frac{\beta_2 \alpha_1}{\alpha_1 - \beta_1}\right) Z_{t1} - \left(\frac{\beta_1 \alpha_2}{\alpha_1 - \beta_1}\right) Z_{t2} + \left(\frac{\beta_3 \alpha_1}{\alpha_1 - \beta_1}\right) Z_{t3} - \left(\frac{\beta_1 \alpha_3}{\alpha_1 - \beta_1}\right) Z_{t4} + \left(\frac{\alpha_1 u_{t1} - \beta_1 u_{t2}}{\alpha_1 - \beta_1}\right)$$

or

$$(11.47) \quad Y_{t1} = \pi_{10} + \pi_{11} Z_{t1} + \pi_{12} Z_{t2} + \pi_{13} Z_{t3} + \pi_{14} Z_{t4} + v_{t1}$$

where

$$\pi_{10} = \frac{\alpha_1 \beta_0 - \beta_1 \alpha_0}{\alpha_1 - \beta_1}$$

$$\pi_{11} = \frac{\beta_2 \alpha_1}{\alpha_1 - \beta_1}$$

$$\pi_{12} = -\frac{\beta_1 \alpha_2}{\alpha_1 - \beta_1}$$

$$\pi_{13} = \frac{\beta_3 \alpha_1}{\alpha_1 - \beta_1}$$

$$\pi_{14} = -\frac{\beta_1 \alpha_3}{\alpha_1 - \beta_1}$$

$$v_{t1} = \frac{\alpha_1 u_{t1} - \beta_1 u_{t2}}{\alpha_1 - \beta_1}$$

The π's are reduced form parameters.

The reduced form equation for price may be obtained by substituting the demand equation of (11.41) for Y_{t1} in (11.45) and solving for price which yields:

$$(11.48)\quad Y_{t2} = \left(\frac{\beta_0 \quad \alpha_0}{\alpha_1 \quad \beta_1}\right) + \left(\frac{\beta_2}{\alpha_1 - \beta_1}\right) Z_{t1} - \left(\frac{\alpha_2}{\alpha_1 - \beta_1}\right) Z_{t2}$$

$$+ \left(\frac{\alpha_3}{\alpha_1 - \beta_1}\right) Z_{t3} - \left(\frac{\alpha 3}{\alpha_1 - \beta_1}\right) Z_{t4} + \left(\frac{u_{t1} - u_{t2}}{\alpha_1 - \beta_1}\right)$$

or

$$(11.49)\quad Y_{t2} = \pi_{20} + \pi_{21} Z_{t1} + \pi_{22} Z_{t2} + \pi_{23} Z_{t3} + \pi_{24} Z_{t4} + v_{t2}$$

where

$$\pi_{20} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$$

$$\pi_{21} = \frac{\beta_2}{\alpha_1 - \beta_1}$$

$$\pi_{22} = -\frac{\alpha_2}{\alpha_1 - \beta_1}$$

$$\pi_{23} = \frac{\beta_3}{\alpha_1 - \beta_1}$$

$$\pi_{24} = -\frac{\alpha_3}{\alpha_1 - \beta_1}$$

$$v_{t2} = \frac{u_{t1} - u_{t2}}{\alpha_1 - \beta_1}$$

The reduced form parameters are those given in (11.47) and (11.49). We have ten expressions and eight unknown structural parameters. Now solving for the structural coefficients we have:

$$(11.50)\quad \beta_1 = \frac{\pi_{12}}{\pi_{22}}$$

but also

$$(11.51)\quad \beta_1 = \frac{\pi_{14}}{\pi_{24}}$$

There is no reason why $\frac{\pi_{12}}{\pi_{22}}$ should equal $\frac{\pi_{14}}{\pi_{24}}$. Hence, there are two ways of obtaining the estimate of β_1 which will yield two estimates of β_1

with no criterion for choosing between the two estimates. In other words in the overidentified case we have more than one way of going from reduced form parameters to structural parameters for some structural parameters. The maintained hypotheses seem to contain "too much" a priori information for some parameters.

The fact that more than one estimate of some structural parameters may be obtained for estimates of reduced form parameters when the structural equations are over-identified is important in terms of estimation. It indicates that estimating the reduced form parameters by least squares or other methods and computing the estimates of structural parameters from these doesn't yield a unique set of estimates of the structural parameters. It tells us that indirect least squares is not an appropriate method of estimation for over-identified equations. When structural equations are over-identified a method of estimation which can utilize all of the a priori information in the maintained hypotheses needs to be used for estimating the structural parameters. Several methods of estimating the parameters of simultaneous systems of equations will be discussed later. One of the more popular of such techniques is two-stage least squares.

Order and Rank Conditions for Identification

The order condition refers to the necessary condition for identification. A necessary condition for a structural equation in a linear model to be identified is that the number of variables, both endogenous and predetermined, excluded from the equation be at least equal to the number of endogenous variables in the system less one. For complete models or systems which contain as many structural equations as endogenous variables then the necessary conditions also is that the number of variables excluded from an equation be at least equal to the number of structural equations less one. For any given structural equation to be just-identified the number of variables excluded from the equation must equal one less than the number of endogenous variables

in the system. If the number of variables excluded from a particular structural equation is greater than one less the number of endogenous variables in the system, the structural equation is over-identified. If the number of variables excluded from a particular structural equation is less than one less the number of endogenous variables in the system the structural equation is under-identified.

To express the order condition symbolically let:

H = total number of predetermined variables in the system.

G = total number of endogenous variables in the system.

h = number of predetermined variables in the particular structural equation being considered.

g = number of endogenous variables in the particular structural equation being considered.

The necessary or order condition for a structural equation to be just-identified is:

$$H + G - (h + g) = G - 1 \qquad (11.51)$$

or

$$H - h = g - 1$$

The latter expression of (11.51) simply indicates that the number of predetermined variables excluded from the structural equation being considered included in the other structural equations of the model has to be one less than the number of endogenous variables in the structural equation being considered.

The necessary or order condition for a structural equation to be over-identified is:

$$H + G - (h + g) > G - 1 \qquad (11.52)$$

$$H - h > g - 1$$

The necessary or order condition for a structural equation to be under-identified is:

$$H + G - (h + g) < g - 1 \qquad (11.53)$$

or

$$H - h < g - 1$$

Order Condition

The order condition for identification is determined for each structural equation prior to estimation. If some structural equations are under-identified then it will not be possible to estimate all of their structural parameters so if estimates of all parameters are needed attempting any estimation will be fruitless. In systems of equations where the parameters of the structural equations may be estimated the estimation procedure to be used will depend on whether the structural equations are just-identified or over-identified. Often times for estimation purposes a distinction is not made between over-identified and just-identified. When a model or structural equations of a model are said to be identified it is understood that they are either just-identified or over-identified.,

The order condition is a necessary condition which has to be fulfilled if identification is to be achieved but just because the order condition is met does not assure us that the model or structural equations of the model are identified. If the order condition is not fulfilled, then we know that identification cannot be achieved but the order condition may be fulfilled and we still may end up with the model and aome or all structural equations not being identified.

Rank Condition

The rank condition is a necessary and sufficient condition for identification. The rank condition for a structural equation is that there exist at least one non-vanishing determinant of rank (G - 1) of the coefficient of the variables, endogenous and predetermined, excluded from the structural equation being considered but appearing in the (G - 1) other structural equations of the model or system. When the rank condition is fulfilled we know that the structural equations with which we are dealing are identified. However, notice that there is a rather subtle difficulty with the rank condition. The difficulty is how one determines which coefficients in the determinat are

non-zero. If we know which parameters in each structural equation were non-zero there would be no problem but we don't know this. After estimation we may be in a position to determine which estimates of parameters may be considered different from zero. So "after the facts" we are in a position to check if the rank condition was fulfilled. Even then there are some difficulties because of a lack of bonefide tests to test hypotheses about structural parameters of equations of a simultaneous system equations.

However, if the assumption is made that all structural parameters of the model are different from zero, before estimation one may determine whether or not it is possible to fulfill the rank condition for each structural equation. This entails finding at least one non-vanishing determinant of rank (G - 1) of the parameters of the variables, endogenous and predetermined, excluded from the structural equation under consideration but appearing in the other (G- 1) structural equations of the model. If such cannot be found for some structural equations then you know that they are not identified even though the order condition may have been fulfilled and any attempt to estimate the structural parameters of the model will be in vain. Furthermore, these determinants provide some insight about which parameters have to be non-zero for the rank condition to be fulfilled and identification achieved. After estimation you can determine whether the parameters required to non-zero in order to fulfill the rank condition can be considered non-zero on the basis of their respective estimates.

Now let us illustrate the order and rank conditions for some models. In all of these models the t subscript of the variables used previously will be dropped. Consider

$$(11.53) \quad Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 Z_1 + \beta_3 Z_2 + u_1$$

$$(11.54) \quad Y_1 = \alpha_0 + \alpha_1 Y_2 + \alpha_2 Z_3 + \alpha_3 Z_4 + u_2$$

The Y's are endogenous and Z's are predetermined. There are two endogenous variables and four predetermined variables; hence H = 4 and G = 2. Checking the order conditions for (11.53) we have (4 - 2) > (2 - 1) and for (11.54) we have (4 - 2) > (2 - 1) hence both are over-identified.

To check the rank condition we have to show that for each equation there is a determinant of rank (G - 1) = (2 - 1) = 1 not equal to zero. Let us set the parameters in a table such as Table 11.1.

TABLE 11.1

Equation	Y_1	Y_2	Z_1	Z_2	Z_3	Z_4
(11.53)	1	β_1	β_2	β_3	0	0
(11.54)	1	α_1	0	0	α_2	α_3

Now for (11.53) the predetermined variables in the system but not in (11.53) are Z_3 and Z_4. Hence if either $\alpha_2 \neq 0$ or $\alpha_3 \neq 0$ then the rank condition for (11.53) is fulfilled. For (11.54) the predetermined variables in the system but not in (11.54) are Z_1 and Z_2. Hence if either $\beta_2 \neq 0$ or $\beta_3 \neq 0$ the rank condition for (11.54) is fulfilled.

Let us consider another illustration of the order and rank conditions utilizing a four equation model. This model could represent the supply and demand equations for two products for which prices and quantities are jointly determined or any system consisting of four relationships and four endogenous variables. Let the model be:

$$\beta_{11} Y_1 + \beta_{12} Y_2 + \beta_{13} Y_3 + \gamma_{11} Z_1 = u_1 \tag{11.55}$$

$$\beta_{21} Y_1 + \beta_{22} Y_2 + \gamma_{22} Z_2 + \gamma_{23} Z_3 = u_2 \tag{11.56}$$

$$\beta_{32} Y_2 + \beta_{33} Y_3 + \beta_{34} Y_4 + \gamma_{31} Z_1 = u_3 \tag{11.57}$$

$$\beta_{43} Y_3 + \beta_{44} Y_4 + \gamma_{42} Z_2 + \gamma_{44} Z_4 = u_4 \tag{11.58}$$

Where Y's are endogenous variables.

Z's are predetermined variables.

β's are parameters of endogenous variables.

γ's are parameters of predetermined variables.

u's are disturbance terms.

The order conditions for (11.55) - (11.58) are:

3 > 2 for (11.55)

2 > 1 for (11.56)

3 > 2 for (11.57)

2 > 1 for (11.58)

All four equations are over-identified in terms of the order condition. For estimation of the structural parameters we would want to choose an estimation procedure appropriate for over-identified systems of simultaneous equations.

Now let us explore the rank condition to determine prior to estimation if it is possible to fulfill the necessary and sufficient condition of identification. In this case there will have to be for each structural equation at least one non-vanishing determinant of rank 3 consisting of the parameters of variables, endogenous and predetermined, not appearing in the equation under considering but appearing in the other 3 structural equations of the model. The required information is presented in Table 11.2.

TABLE 11.2

Equation	Y_1	Y_2	Y_3	Y_4	Z_1	Z_2	Z_3	Z_4
(11.55)	β_{11}	β_{12}	β_{13}	0	γ_{11}	0	0	0
(11.56)	β_{21}	β_{22}	0	0	0	γ_{22}	γ_{23}	0
(11.57)	0	β_{32}	β_{33}	β_{34}	γ_{31}	0	0	0
(11.58)	0	0	β_{43}	β_{44}	0	β_{42}	0	γ_{44}

For (11.55) we have the following submatrix of parameters to consider

$$(11.59)\quad \begin{bmatrix} 0 & \gamma_{22} & \gamma_{23} & 0 \\ \beta_{34} & 0 & 0 & 0 \\ \beta_{44} & \gamma_{42} & 0 & \gamma_{44} \end{bmatrix}$$

from which the following four 3 x 3 determinants may be formed:

$$\begin{vmatrix} 0 & \gamma_{22} & \gamma_{23} \\ \beta_{34} & 0 & 0 \\ \beta_{44} & \gamma_{42} & 0 \end{vmatrix} = \gamma_{23}\ \gamma_{42}\ \beta_{34}$$

$$\begin{vmatrix} 0 & \gamma_{22} & 0 \\ \beta_{34} & 0 & 0 \\ \beta_{44} & \gamma_{42} & \gamma_{44} \end{vmatrix} = -\gamma_{22}\ \beta_{34}\ \gamma_{44}$$

$$\begin{vmatrix} 0 & \gamma_{23} & 0 \\ \beta_{34} & 0 & 0 \\ \beta_{44} & 0 & \gamma_{44} \end{vmatrix} = -\gamma_{23}\ \beta_{34}\ \gamma_{44}$$

$$\begin{vmatrix} \gamma_{22} & \gamma_{23} & 0 \\ 0 & 0 & 0 \\ \gamma_{42} & 0 & \gamma_{44} \end{vmatrix} = 0$$

Three of these four determinants may be non-zero if the parameters involved are not zero. However, note that all three involve β_{34} and $\beta_{34} = 0$ all determinants of order 3 for (11.55) are zero.

For (4.56) we have the following submatrix of parameters to consider:

$$(11.60)\quad \begin{bmatrix} \beta_{13} & 0 & \gamma_{11} & 0 \\ \beta_{33} & \gamma_{34} & \gamma_{31} & 0 \\ \beta_{43} & \beta_{44} & 0 & \gamma_{44} \end{bmatrix}$$

from which the following four 3 x 3 determinants may be formed.

$$\begin{vmatrix} \beta & 0 & \gamma_{11} \\ \beta_{33} & \beta_{34} & \gamma_{31} \\ \beta_{43} & \beta_{44} & 0 \end{vmatrix} = \gamma_{11}\ \beta_{33}\ \beta_{44} - \gamma_{11}\ \beta_{34}\ \beta_{43} - \gamma_{31}\ \beta_{44}\ \beta_{18}$$

$$\begin{vmatrix} \beta_{13} & 0 & 0 \\ \beta_{33} & \beta_{34} & 0 \\ \beta_{43} & \beta_{44} & \gamma_{44} \end{vmatrix} = \beta_{13}\ \beta_{34}\ \gamma_{44}$$

$$\begin{vmatrix} \beta_{13} & \gamma_{11} & 0 \\ \beta_{33} & \gamma_{31} & 0 \\ \beta_{43} & 0 & \gamma_{44} \end{vmatrix} = \beta_{13}\ \gamma_{31}\ \gamma_{44} - \gamma_{11}\ \beta_{33}\ \gamma_{44}$$

$$\begin{vmatrix} 0 & \gamma_{11} & 0 \\ \beta_{34} & \gamma_{31} & 0 \\ \beta_{44} & 0 & \gamma_{44} \end{vmatrix} = -\gamma_{11}\ \beta_{34}\ \gamma_{44}$$

Note that if $\gamma_{11} = 0$ and $\beta_{13} = 0$ that the rank condition for (11.56) would not be fulfilled even if the other parameters were non zero.

For (11.57) we have the following submatrix of parameters to consider:

(11.61)
$$\begin{vmatrix} \beta_{11} & 0 & 0 & 0 \\ \beta_{21} & \gamma_{22} & \gamma_{23} & 0 \\ 0 & \gamma_{42} & 0 & \gamma_{44} \end{vmatrix}$$

from which the following four 3 x 3 determinants may be found:

$$\begin{vmatrix} \beta_{11} & 0 & 0 \\ \beta_{21} & \gamma_{22} & \gamma_{23} \\ 0 & \gamma_{42} & 0 \end{vmatrix} = \beta_{11}\ \gamma_{42}\ \gamma_{23}$$

$$\begin{vmatrix} \beta_{11} & 0 & 0 \\ \beta_{21} & \gamma_{22} & 0 \\ 0 & \gamma_{42} & \gamma_{44} \end{vmatrix} = \beta_{11}\ \gamma_{22}\ \gamma_{44}$$

$$\begin{vmatrix} \beta_{11} & 0 & 0 \\ \beta_{21} & \gamma_{23} & 0 \\ 0 & 0 & \gamma_{44} \end{vmatrix} = \beta_{11}\ \gamma_{23}\ \gamma_{44}$$

$$\begin{vmatrix} 0 & 0 & 0 \\ \gamma_{22} & \gamma_{23} & 0 \\ \gamma_{42} & 0 & \gamma_{44} \end{vmatrix} = 0$$

Note that if $\beta_{11} = 0$ then all of the determinants of order 3 are zero and the rank condition for (11.57) would not be fulfilled.

For (11.58) we have the following submatrix of parameters to consider:

$$\left[\begin{array}{ccc|c} \beta_{11} & \beta_{12} & \gamma_{11} & 0 \\ \beta_{21} & \beta_{22} & 0 & \gamma_{23} \\ 0 & \beta_{32} & \gamma_{31} & 0 \end{array}\right]$$

from which the following four 3 x 3 determinants may be formed:

$$\begin{vmatrix} \beta_{11} & \beta_{12} & \gamma_{11} \\ \beta_{21} & \beta_{22} & 0 \\ 0 & \beta_{32} & \gamma_{31} \end{vmatrix} = \beta_{11}\beta_{22}\gamma_{31} + \gamma_{11}\beta_{21}\beta_{32} - \beta_{12}\beta_{21}\gamma_{31}$$

$$\begin{vmatrix} \beta_{11} & \beta_{12} & 0 \\ \beta_{21} & \beta_{22} & \gamma_{23} \\ 0 & \beta_{32} & 0 \end{vmatrix} = -\beta_{11}\beta_{32}\gamma_{23}$$

$$\begin{vmatrix} \beta_{11} & \gamma_{11} & 0 \\ \beta_{21} & 0 & \gamma_{23} \\ 0 & \gamma_{31} & 0 \end{vmatrix} = -\beta_{11}\,\gamma_{31}\,\gamma_{23}$$

$$\begin{vmatrix} \beta_{12} & \gamma_{11} & 0 \\ \beta_{22} & 0 & \gamma_{23} \\ \beta_{32} & \gamma_{31} & 0 \end{vmatrix} = \gamma_{11}\,\gamma_{23}\,\beta_{32} - \beta_{12}\,\gamma_{31}\,\gamma_{23}$$

All four of the determinants are candidates for fulfilling the rank condition and there is no one parameter if it were zero that would make all determinants of order 3 to vanish as was the case for (11.57).

Other Information Which May Facilitate Identification

In the discussion of various forms of identification focus has been primarily on various amounts of a priori information in the maintained hypotheses and restrictions on parameters in the structural relationships. The a priori information has been in terms of predetermined variables in various relationships and in light of restrictions that their respective parameters were not zero. Information about the characteristics of the disturbances of structural equations may also be helpful in achieving identification, often in relative sense, of certain structural equations. The uses in identification of various types of information about disturbances are disnussed briefly. The purpose is to draw attention to the usefulness of such information, if available, in identifying structural equations. One of the major shortcomings of using information about the characteristics of disturbances in achieving identification is that very seldomly is such information available.

Independence of Disturbances

If the disturbances of the structural equations of a model are known or assumed, on one basis or another, to be independent of each other then there are no problems with identification of the structural relationships. Independence of disturbances of the various structural relationships indicate that the forces which operate within each of the structures are independent and thus unrelated to the forces which operate in any one of the other structures in the model. This means that the system of equations is not simultaneous and that the parameters of each structural equation may be estimated by considering each structural equation separately. Least squares or other single equation methods of estimation could be used to estimate the parameters of the individual structural equations. This basically is the underlying rationale for recursive systems of equations and the independence of disturbances is based on assumption.

Uncorrelated Disturbances

The implications of uncorrelated disturbances are similar to the implications of independent disturbances. Uncorrelated disturbances are a weaker restriction than independent disturbances. However, in both cases we are talking about unrelated disturbances. Uncorrelated disturbances refer to the lack of a linear association among the disturbances of the structural equations based on the particular sample of disturbances being considered. This does not mean that the populations of the true disturbances are in fact independent. However, independence of disturbances among structural equations does imply uncorrelated disturbances among structural equations.

In the case of uncorrelated disturbances the forces operating in a structure are unrelated to the forces operating in the other structures of the model. In terms of identification of the structural equations and estimation of their parameters the situation is the same as for independent disturbances.

Limit To The Size of Disturbance

Suppose we have adequate prior information about the magnitudes of the disturbance of one or more of the structural equations of the model. Let the maximum deviation of a disturbance of a structural equation be:

(11.63) $\left|u_{tj}\right| \leq c$

where

u_{tj} is the disturbance for the jth equation in time period t.

c is some constant value.

If $c = 0$ then $u_{tj} = 0$ which indicates that the observations are on the curve representing the jth equation and the jth equation is identified.

If $c \neq 0$ then (11.63) provides information about the range of the disturbance of the jth equation whatever that equation might be representing. Based on (11.63) the range would be 2c which would establish bounds for the observations and this information combined with the nature of the scatter of observations might help us identify one or more structures of the model. For example suppose the range on u_{tj} and the scatter of data were as in Figure 11.8. Suppose Y_2 represents price

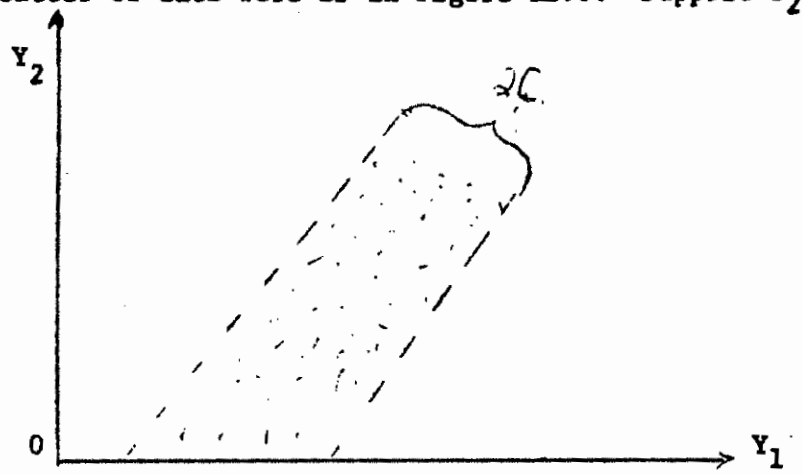


Figure 11.8

and Y_1 quantity. If this were a supply-demand phenomena then the range of u_{tj} would identify the jth equation and in this case it would be a supply equation. The demand equation would not be identified on the basis of this information. On the other hand, if the range of u_{tj} and scatter of data were as in Figure 11.9, the jth equation would be identified and it would be a demand equation.

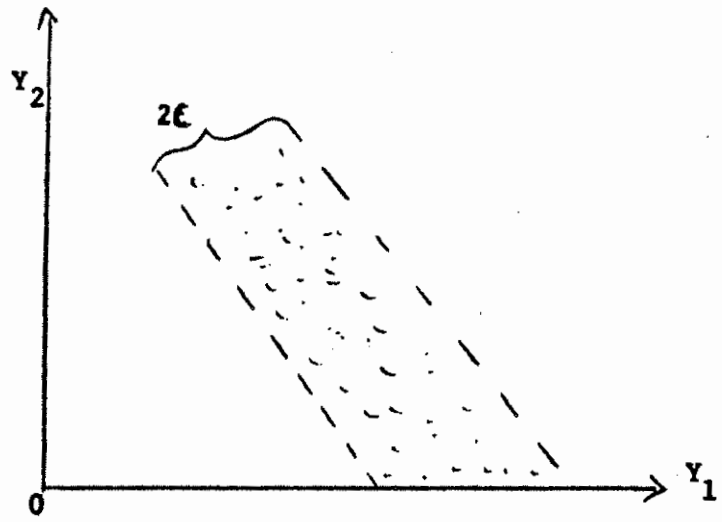


Figure 11.9

Variance of One Disturbance Less Than the Variance of the Other.

Again let us consider the model:

(11.64) D: $Y_{t1} = \beta_0 + \beta_1 Y_{t2} + u_{t1}$

S: $Y_{t1} = \alpha_0 + \alpha_1 Y_{t2} + \alpha_2 Z_{t1} + u_{t2}$

where Y_{t1} is endogenous and represents quantity of t.

Y_{t2} is endogenous and represents price of t.

Z_{t1} is a predetermined variable in t.

u_{t1} is the disturbance of the demand equation.

u_{t2} is the disturbance of the supply equation.

β's are parameters of the demand equation.

α's are parameters of the supply equation.

Suppose we have a priori information that the variance of u_{t2} is considerably less than the variance of u_{t1} i.e., $\sigma^2_{u_2} < \sigma^2_u$.

In this case the demand equation would be identified because of the predetermined variable, Z_{t1}, in the supply equation, but which does not appear in the demand equation. But if the pattern of observations on Y_1 and Y_2 is as in Figure 11.10 the supply equation is also identified. Note that for any given

level of X_{t1} the supply curve is shifting only a little while the demand curve shifting considerably. Knowing that the variance of the disturbance of the supply equation is considerably less than the variance of the demand equation allows approximate identification of the slope parameter of the supply equation.

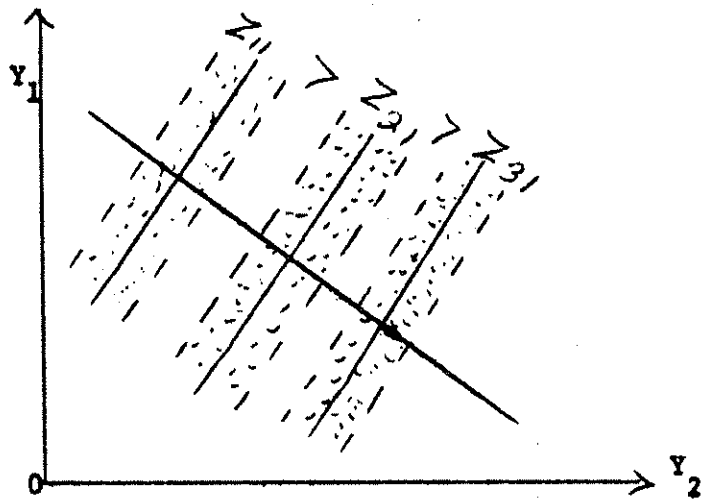


Figure 11.10

CHAPTER 12

Introduction to Simultaneous Linear Systems of Equations

Introduction

There are two general methods of estimation available for estimating the parameters of identified simultaneous linear systems of equations - full information methods and limited information methods. Limited information methods pertain to a sub-set of the system of equations comprising a model. The sub-set may be and often is a single equation. When the sub-set is a single equation the estimation is performed by considering one equation at a time.

Full Information Methods

Full information methods utilize all the restrictions of the structural equations. This means information about the endogenous and predetermined variables included in each equation and excluded from each equation but in the remainder of the model (in each equation which parameters are zero and which are not) and the specific algebraic expression of each equation in the model is utilized in full information methods. Hence, to use full information methods of estimation the algebraic form and the specific variables, variables in each and every equation have to be specified. All parameters of the entire model are then estimated concurrently. As might be expected the computational procedures for full information methods are considerably more complex than limited information methods. The two common full information methods are full information maximum likelihood (or full information least generalized residual variance) method and three stage least squares. The computational task for three stage least squares is considerably less than for full information maximum likelihood.

Limited Information Methods

Limited information methods of estimation of parameters of structural equations of a linear simultaneous system of equation utilizes only part of the restrictions of the structural equations. The method uses only information derived from knowledge of the restrictions pertaining to the sub-set of equations which are of interest. One of the most commonly used sub-sets is a single structural equation. There is little or no advantage to using sub-sets of more thsn one equation and some of the limited information methods apply to sub-sets which are a single structural equation. The limited information methods of estimation pertaining to single structural equations are usually referred to as single equation methods. Generally the computational task of single equation methods is considerably less burdensome than those of multiple structural equation subsets.

Limited information methods allow estimating the parameters of part of the system of structural equations on the basis of a minimum amount of knowledge about the entire system. The part of the system may consist of one structural equation, any number of all structural equations of the system. In single equation methods of estimation the information about the restrictions of the structural equations is utilized on a single equation basis. In single equation methods information about all endogenous and predetermined variables in the entire aystem is required. The specific algebraic form of the structural equation of interest has to be specified and the endogenous and predetermined variables included in that particular structural equation and those excluded from that particular structural equation but in the other structural equations of the model have to be known and specified. In other words in terms of a priori information about parameters we have to have information about which parameters of the particular structural equation of interest are non-zero and which are zero and which parameters of the other structural equations are non-zero. But with respect to the variables excluded from the

particular structural equation of interest we do not need to know in which particular other structural equation the excluded endogenous and predetermined variable appear. Nor do we need to know the specific formulations and algebraic forms of the other structural equations of the model. Hence, we can estimate the parameters of a particular structural equation, the algebraic form of the particular structural equation, and the variables included in the remaining sub-set of structural equations.

There are a number of single equation methods for estimating parameters of structural equations available. We will concern ourselves mainly with four such methods. The two methods which we will consider in greatest detail are two stage least squares and limited information maximum likelihood (or limited information least variance ratio). Two stage least squares is sometimes also referred to as the Theil-Basmann method. It is one of the most commonly used methods of estimation of parameters of an over-identified system of simultaneous equations. The method of instrumental variables and Theil's k-class estimators will also be considered.

Properties of Estimators

Before investigating methods of estimating structural parameters of simultaneous systems it might be helpful to briefly mention some of the general properties of estimators of structural parameters of simultaneous equation systems.[1/] First all of the estimators mentioned - two stage least squares, limited information maximum likelihood, k-class, instrumental variables, three stage least squares, and full information maximum likelihood are biased. But if the disturbances are serially independent the methods of estimation yield consistent estimators. The k-class estimators are consistent if the probability limit of (k-1) = 0.

[1/] For a summary of properties of several methods for structures and for reduced forms see Carl Christ, Econometric Methods and Models, pp. 464-481 and particularly Table 12.1 on page 466. A summary of Monte Carlo studies of properties of estimators of parameters of simultaneous systems of equations is presented in Chapter 13 of J. Johnston, Econometric Methods, Second Edition, McGraw-Hill, 1972.

For large sample sizes the limited information maximum likelihood and full information maximum likelihood methods yield estimators which are asympotically normal if disturbances are serially independent. Asympotic covariances matrices are available for two stage least squares and three stage least squares estimators. Approximate covariance matrices are available for limited information and k-class estimators. If the disturbances are normally distributed then all the methods of estimation mentioned except the instrumental variable method yield asymptotically efficient estimators where asymtotic efficiency for limited information maximum likelihood, two stage least squares, and k-class estimators refers to asymptotic efficiency relative to other estimators using the same incomplete a priori information. Overall the full information methods generally yield more efficient estimators than limited information methods. However, from an applied viewpoint the efficiency of full information estimators may not be of any significant consequences.

For just-identified simultaneous systems of equations, two stage least squares, limited information maximum likelihood, instrumental variables, and indirect least squares all yield identical estimators. Indirect least squares entails estimating the parameters of reduced form equations by ordinary least squares methods and computing the estimates of structural parameters from the estimated reduced form coefficients.

Normalization of Equations

Normalization of an equation entails making the parameter of one of the endogenous variables one. This facilitates distinguishing among the structures of the model and for some estimation methods is a necessary step in the estimation process. The choice of the endogenous variable with respect to which a particular structural equation is normalized is arbitrary, however, depending upon the phenomena represented by the

structure normalization with respect to a particular endogenous variable may make more economic sense and economic analysis may be more amenable than if normalization were with respect to some other endogenous variable. The normalization of equations is generally taken into consideration, implicitly or explicitly in model formulation and the variables' included in a particular structural equation are in part chosen on the basis of the particular normalization. For example, if a demand equation is normalized with respect to quantity then very likely the prices of substitute products will be included in the equation whereas if the structure is normalized with respect to price then very likely quantities of substitute products will be included. However, the mathematical (really mechanical) aspects of normalization are strictly arbitrary.

For illustrative purposes let us consider a rather simple structural equation from some system of simultaneous equations. Let the system be:

$$\beta Y + \Gamma Z = U. \tag{12.1}$$

where no signs are being imposed on the disturbances U. Now suppose we select one equation out of this system, the ith one, and it is of the form

$$\beta_{i1} Y_{i1} + \beta_{i2} Y_{i2} + \gamma_{i1} Z_{i1} + \gamma_{i2} Z_{i2} = u_i \tag{12.2}$$

We can multiply this equation by a non-zero constant and not really change the content of the equation. But first let us choose to normalize this equation with respect to Y_{i1} so

$$Y_{i1} + \frac{\beta_{i2}}{\beta_{i1}} Y_{i2} + \frac{\gamma_{i1}}{\beta_{i1}} Z_{i1} + \frac{\gamma_{i2}}{\beta_{i1}} Z_{i2} = \frac{u_i}{\beta_{i1}} \tag{12.3}$$

Now set $\beta_{i1} = 1$ and

$$Y_{i1} + \beta_{i2} Y_{i2} + \gamma_{i1} Z_{i1} + \gamma_{i2} Z_{i2} = u_i \tag{12.4}$$

where (12.4) is (12.2) normalized with respect to the Y_{i1} endogenous variable.

The other structural equations of (12.1) may be normalized in a similar manner. In each equation the coefficient of one endogenous variable is set equal to some non-zero constant usually one or minus one depending on the particular formulation of the model.

Note that if the normalization would have been done by letting $\beta_{i1} = -1$ then the normalized equation would be:

(12.5) $$Y_{i1} - \beta_{i2} Y_{i2} - \gamma_{i1} Z_{i1} - \gamma_{i2} Z_{i2} = -u_i$$

or

(12.6) $$Y_{i1} = \beta_{i2} Y_{i2} + \gamma_{i1} Z_{i1} + \gamma_{i2} Z_{i2} - u_i$$

where the sign on the u_i is immaterial since whether the disturbance is incorporated into an equation with positive or negative sign is of no consequence.

If the model had been expressed in implicit form

(12.7) $$\beta Y + \Gamma Z + U = 0$$

then the ith equation would be

(12.8) $$\beta_{i1} Y_{i1} + \beta_{i2} Y_{i2} + \gamma_{i1} Z_{i1} + \gamma_{i2} Z_{i2} + u_i = 0$$

normalizing by setting $\beta_{i1} = -1$ yields

(12.9) $$Y_{i1} - \beta_{i2} Y_{i2} - \gamma_{i1} Z_{i1} - \gamma_{i2} Z_{i2} - u_i = 0$$

and

(12.10) $$Y_{i1} = \beta_{i2} Y_{i2} + \gamma_{i1} Z_{i1} + \gamma_{i2} Z_{i2} + u_i$$

Regardless of which of these normalization schemes are used as long as the signs are considered appropriately the final estimated equations will be the same.

Least Squares Applied to Structural Equations

When the joint determination of variables in a structural equation are disregarded and ordinary least squares methods are used to estimate the parameters of a structural equation the estimators are not only biased but they are inconsistent. Hence, the bias persists even for large sample sizes. This phenomena is referred to as "least squares bias." This should not be misinterpreted that the least squared method is biased but simply that ordinary least squares yields biased and inconsistent estimators under these conditions.

The reason for this bias and inconsistency is that in a structural equation from a simultaneous equation model there is more than one endogenous variable per equation and hence some of the regressors are interdependent with the disturbance of the structural equation. To explore the basis of the bias and inconsistency let us consider the following just-identified two equations model:

$$\begin{aligned} \beta_{11} Y_1 + \beta_{12} Y_2 + \gamma_{11} Z_1 + \gamma_{10} Z_0 &= u_1 \\ \beta_{21} Y_1 + \beta_{22} Y_2 + \gamma_{22} Z_2 + \gamma_{20} Z_0 &= u_2 \end{aligned} \tag{12.11}$$

Normalizing the first equation with respect to Y_1 and the second with respect to Y_2 by setting their respective parameters equal to minus one yields

$$\begin{aligned} Y_1 - \beta_{12} Y_2 - \gamma_{11} Z_1 - \gamma_{10} Z_0 &= -u_1 \\ -\beta_{21} Y + Y_2 - \gamma_{22} Z_2 - \gamma_{20} Z_0 &= -u_2 \end{aligned} \tag{12.12}$$

or

$$\begin{aligned} Y_1 &= \beta_{12} Y_2 + \gamma_{11} Z_1 + \gamma_{10} Z_0 - u_1 \\ Y_2 &= \beta_{21} Y_1 + \gamma_{22} Z_2 + \gamma_{20} Z_0 - u_2 \end{aligned} \tag{12.13}$$

The reduced form equations for the endogenous variables are:

$$Y_1 = \frac{(\beta_{12}\gamma_{20} + \gamma_{10})}{(1 - \beta_{12}\beta_{21})} Z_0 + \frac{\gamma_{11}}{(1 - \beta_{12}\beta_{21})} Z_1$$

(12.14)

$$+ \frac{\beta_{12}\gamma_{22}}{(1 - \beta_{12}\beta_{21})} Z_2 - \frac{(u_1 + \beta_{12}u_2)}{(1 - \beta_{12}\beta_{21})}$$

$$Y_2 = \frac{(\beta_{21}\gamma_{10} + \gamma_{20})}{(1 - \beta_{12}\beta_{21})} Z_0 + \frac{\beta_{21}\beta_{11}}{(1 - \beta_{12}\beta_{21})} Z_1$$

$$+ \frac{\gamma_{22}}{(1 - \beta_{12}\beta_{21})} Z_2 - \frac{(\beta_{21}u_1 + u_2)}{(1 - \beta_{12}\beta_{21}}$$

Now suppose we obtain estimates of the parameters of the first equation of (12.13) by applying ordinary least squares directly to the structural equation. Let us consider a sample size of n observations and to facilitate the illustration rewrite the equation in matrix form as follows:

$$(12.15) \quad Y_1 = [Y_2 Z_1 Z_0] \begin{bmatrix} \beta_{12} \\ \gamma_{11} \\ \gamma_{10} \end{bmatrix} - u_1$$

where Y_1, Y_2, Z_1, Z_0, and u_1 are all n x 1 vectors. Now let

$$Z = [Z_1 \; Z_0] \quad \text{and} \quad \gamma = \begin{bmatrix} \gamma_{11} \\ \gamma_{10} \end{bmatrix}$$

and the equation may be expressed as

$$(12.16) \quad Y_1 = [Y_2 \; Z] \begin{bmatrix} \beta_{12} \\ \gamma \end{bmatrix} - u_1$$

The least squares estimators of the parameters are:

$$(12.17)\quad \begin{bmatrix} \hat{\beta}_2 \\ \hat{\gamma} \end{bmatrix} = \{[Y_2\ Z]'\ [Y_2\ Z]\}^{-1}\ [Y_2\ Z]'\ Y_1$$

$$= \begin{bmatrix} Y_2'\ Y_2 & Y_2'\ Z \\ Z'\ Y_2 & Z'Z \end{bmatrix}^{-1} [Y_2 Z]'\ Y_1$$

$$= \begin{bmatrix} Y_2'\ Y_2 & Y_2'\ Z \\ Z'\ Y_2 & Z'Z \end{bmatrix}^{-1} [Y_2 Z]'\ [Y_2 Z] \begin{bmatrix} \beta_{12} \\ \gamma \end{bmatrix}$$

$$- \begin{bmatrix} Y_2'\ Y_2 & Y_2'\ Z \\ Z'\ Y_2 & Z'Z \end{bmatrix}^{-1} [Y_2 Z]'\ u_1$$

$$= \begin{bmatrix} \beta_{12} \\ \gamma \end{bmatrix} - \begin{bmatrix} Y_2'\ Y_2 & Y_2'\ Z \\ Z'\ Y_2 & Z'Z \end{bmatrix}^{-1} \begin{bmatrix} Y_2'\ u_1 \\ Z'\ u_1 \end{bmatrix}$$

The expected value of the second component of the last expression of (12.17) is not zero nor does the component vanish if we take the probability limit.[2/] We know from the formulation of the model that Y_2 is not uncorrelated with the disturbance, u_1. Furthermore, $Y_2'\ Y_2$ of the inverse matrix

$$(12.18)\quad \begin{bmatrix} Y_2'\ Y_2 & Y_2'\ Z \\ Z'\ Y_2 & Z'Z \end{bmatrix}^{-1}$$

[2/] If the mass or density function of a sequence of random variables $\{x^{(n)}\}$ becomes entirely concentrated at some point as n goes to infinity, the sequence is said to converge in probability to that point and that point is said to be the probability limit of the sequence. If the entire distribution collapses on a point that point is the asymptotic expectation. However, an expectation can approach a constant without the distribution collapsing. If the variance goes to zero the distribution will collapse. See Arthur S. Goldberger, Econometric Theory, John Wiley, 1964, Chapter 6.

will result in variance and covariance terms of the disturbances. Hence, the estimators are biased.

Possibly we can gain some further insight about this bias if we express the reduced form for Y_2 of (12.14) in terms of reduced form parameters and substitute for Y_2 in (12.17). Let

$$(12.19)\quad Y_2 = \pi_{20} + \pi_{21} Z_1 + \pi_{22} Z_2 + v_2$$

where $$\pi_{20} = \frac{\beta_{21}\gamma_{10} + \gamma_{20}}{1 - \beta_{12}\beta_{21}}$$

$$\pi_{21} = \frac{\beta_{21}\gamma_{11}}{1 - \beta_{12}\beta_{21}}$$

$$\pi_{22} = \frac{\gamma_{22}}{1 - \beta_{12}\beta_{21}}$$

$$v_2 = \frac{(\beta_{21}u_1 + u_2)}{1 - \beta_{12}\beta_{21}}$$

and

$$(12.20)\quad Y_2 = [Z_0 Z_1 Z_2] \begin{bmatrix} \pi_{20} \\ \pi_{21} \\ \pi_{22} \end{bmatrix} + v_2 = [Z_1\ Z_0\ Z_2] \begin{bmatrix} \pi_{21} \\ \pi_{20} \\ \pi_{22} \end{bmatrix} + v$$

but $Z = [Z_1 Z_0]$ as defined for (12.16) so if we let $\pi = [\pi_{21}\ \pi_{20}]'$ then

$$(12.21)\quad Y_2 = [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} + v_2$$

Now

$$(12.22)\quad Y_2'\,Y_2 = \left([Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} + v_2\right)' \left([Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} + v_2\right)$$

$$= \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}' [Z\ Z_2]' [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} + v_2' [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} + \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}' [Z\ Z_2]' v_2 + v_2' v_2$$

$$= \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}' [Z\ Z_2]' [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} + \frac{1}{(1 - \beta_{12}\beta_{21})} [\beta_{21}u_1+u_2]' [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}$$

$$+ \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}' [Z\ Z_2]' \frac{1}{(1-\beta_{12}\beta_{21})} [\beta_{21}u_1+u_2] + \frac{1}{(1-\beta_{12}\beta_{21})^2} [\beta_{21}u_1+u_2]' [\beta_{21}u_1+u_2]$$

and

$$(12.23)\quad Y_2' Z = \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}' [Z\ Z_2]' Z - \frac{(\beta_{21} u_1 + u_2)'}{1 - \beta_{12}\beta_{21}} Z$$

and

$$(12.24)\quad Z'Y_2 = Z' [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} - Z' \frac{(\beta_{21} u_1 + u_2)}{1 - \beta_{12}\beta_{21}}$$

and

$$(12.25)\quad Y_2' u_1 = \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} [Z\ Z_2]' u_1 - \frac{(\beta_{21} u_1 + u_2)'}{1 - \beta_{12}\beta_{21}} u_1$$

Now the probability limits of (12.22) - (12.25) are

$$(12.26)\quad \operatorname*{plim}_{n\to\infty} \frac{Y_2' Y_2}{n} = \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}' [Z\ Z_2]' [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} + \frac{\beta_{21}^2 \sigma_1^2 + 2\beta_{21}\sigma_{12}+\sigma_2^2}{(1 - \beta_{12}\beta_{21})^2}$$

$$(12.27)\quad \operatorname*{plim}_{n\to\infty} \frac{Y_2' Z}{n} = \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}' [Z\ Z_2]' Z$$

$$(12.28)\quad \operatorname*{plim}_{n\to\infty} \frac{Z' Y_2}{n} = Z' [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}$$

$$(12.29)\quad \operatorname*{plim}_{n\to\infty} \frac{Y_2' u_1}{n} = \frac{-\beta_{21}\sigma_1^2-\sigma_{12}}{1 - \beta_{12}\beta_{21}}$$

and hence the probability limits of the estimators are:

(12.30)

$$\operatorname*{plim}_{n \to \infty} \begin{bmatrix} \hat{\beta}_{12} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \beta_{12} \\ \gamma \end{bmatrix} - \left[\left\{ \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} [Z\ Z_2]' [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} + \frac{\beta_{21}^2 \sigma_1^2 + 2\beta_{21}\sigma_{12} + \sigma_2^2}{(1 - \beta_{12}\beta_{21})^2} \left\{ \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}' [Z\ Z_2]' Z \right\} \right\} \right.$$

$$\left. \left\{ Z' [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} \right\} \left\{ Z'Z \right\} \right]^{-1} \cdot \begin{bmatrix} \frac{-\beta_{21}\sigma_1^2 - \sigma_{12}}{1 - \beta_{12}\beta_{21}} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \beta_{12} \\ \gamma \end{bmatrix} + \left[\left\{ \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}' [Z\ Z_2]' [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} + \frac{\beta_{21}^2 \sigma_1^2 + 2\beta_{21}\sigma_{12} + \sigma_2^2}{(1 - \beta_{12}\beta_{21})^2} \left\{ \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix}' [Z\ Z_2]' Z \right\} \right\} \right.$$

$$\left. \left\{ Z' [Z\ Z_2] \begin{bmatrix} \pi \\ \pi_{22} \end{bmatrix} \right\} \left\{ Z'Z \right\} \right]^{-1} \cdot \begin{bmatrix} \frac{\beta_{21}\sigma_1^2 + \sigma_{12}}{1 - \beta_{12}\beta_{21}} \\ 0 \end{bmatrix}$$

From (12.30) it is clear that the bias does not vanish as $n \to \infty$ and hence the estimators are not consistent. The bias terms are functions of the variances and covariances of the disturbances and the structural parameters.

Although only one structural equation was considered in illustrating the bias and inconsistency of ordinary least squares when applied to a structural equation from a simultaneous equation model, the bias and inconsistency of least squares estimators could be developed for the other equations in the model.[3/] Also, the model considered here was a two equation just-identified model but the same would hold true for models consisting of more than two equations and for over-identified models.[4/]

[3/] For an illustration of "least squares bias' for model consisting of a stochastic consumption function and a non-stochastic income relation see Christ pages 459-464 or Johnston pages 341-346.

[4/] See Jean Bronfenbrenner "Source and Size of Least-Squares Bias in a Two-Equation Model", in Wm. C. Hood and Tjalling C. Koopmans. Studies Chapter IX, Pages 221-235.

CHAPTER 13

Two Stage Least Squares Estimation

Introduction

Two-stage least squares (TSLS) is one of the most commonly used limited information single equation methods for estimating the parameters of a simultaneous system of linear equations. In terms of the estimation process itself, it is one of the more easily understood methods of estimation of parameters of simultaneous systems of linear equations. Also, the computational task is one of the least burdensome and it is one of the lower cost computational methods. Furthermore, based on Monte Carlo studies the sample properties of TSLS estimators are about as good or at least not inferior to other estimation methods.[1/]

First a simultaneous equation model in general matrix form will be specified and the underlying assumptions indicated. Then we will take an intuitive and "cookbook" look at TSLS. This will be followed by a rather detailed development of the TSLS method of estimation.

The Model and Assumptions

The type of model we will be considering is:

$$(13.1) \quad Y\beta + Z\Gamma = U$$

where Y is a (T x M) matrix of observations of endogenous variables.
β is a (M x M) matrix of parameters of the endogenous variables.
Z is a (T x N) matrix of observations of predetermined variables.
Γ is a (N x M) matrix of parameters of the predetermined variables.
U is a (T x M) matrix of disturbances.

1/ See J. G. Cragg, "On the Relative Small-Sample Properties of Several Structural Equation Estimators," Econometrica, Vol. 35, No. 1, January 1967, pp. 89-110, J. Johnston, Econometric Methods, McGraw Hill, 1972, pp. 408-420, and Jan Kmenta, Elements of Econometrics, Macmillan, 1971, pp. 581-585.

There are M endogenous variables in the simultaneous system of equations and M equations and in this sense the model is complete. There are N predetermined variables in the entire model and of course if the structural parameters are to be estimated the predetermined variables will have to be distributed among the structural equations in such a manner that the structural equations are identified.

We will be making the following assumptions:

1. Model is linear in parameters.
2. Model's determinant is not zero (determinant of β matrix not zero).
3. Z's are predetermined and linearly independent.

(13.2)

4. Equations are identifiable (just-identified or over-identified).
5. u's are random with zero mean and finite variances and covariances.

$$E(u_i u_j) = \emptyset$$

6. Distribution of u's is the same in all time periods; their variance-covariance matrix is non-singular.
7. u's are serially non-correlated;
8. u's are normally distributed.
9. Moment matrix of predetermined variables is well behaved in the limit (in the limit Z's exist and is non-singular)

The last two assumptions facilitate the designation of asymptotic properties of the estimators.

A Cursory Look at Two Stage Least Squares

As its name implies TSLS entails applying least squares procedures in two stages. To illustrate the procedure in somewhat of a "cookbook" fashion, let us consider a two-equation over-identified simultaneous equation model which will be assumed to fulfill the assumption specified in (13.2). The model to be considered is:

$$\beta_{11} Y_1 + \beta_{12} Y_2 + \gamma_{11} Z_1 + \gamma_{12} Z_2 + \gamma_{13} Z_3 + \gamma_{10} Z_0 + u_1 = 0$$

(13.3)

$$\beta_{21} Y_1 + \beta_{22} Y_2 + \gamma_{21} Z_1 + \gamma_{24} Z_4 + \gamma_{25} Z_5 + \gamma_{20} Z_0 + u_2 = 0$$

Let us normalize the first equation of (13.3) with respect to Y_1 and the second equation with respect to Y_2 such that the normalized equations ($\beta_{11} = \beta_{22} = -1$) are as follows:

$$Y_1 = \beta_{12} Y_2 + \gamma_{11} Z_1 + \gamma_{12} Z_2 + \gamma_{13} Z_3 + \gamma_{10} Z_0 + u_1$$

(13.4)

$$Y_2 = \beta_{21} Y_1 + \gamma_{21} Z_1 + \gamma_{24} Z_4 + \gamma_{25} Z_5 + \gamma_{20} Z_0 + u_2$$

where in both (13.3) and (13.4)

Y's are endogenous variables.

Z's are predetermined variables with Z_0 a vector of ones.

β's are the parameters of endogenous variables.

γ's are the parameters of predetermined variables.

u's are vectors of disturbances fulfilling 5, 6, 7, and 8 of (13.2)

The first stage equations of TSLS, which are a type of reduced form equations, are:

$$Y_1 = \pi_{10} Z_0 + \pi_{11} Z_1 + \pi_{12} Z_2 + \pi_{13} Z_3 + \pi_{14} Z_4 + \pi_{15} Z_5 + v_1$$

(13.5)

$$Y_2 = \pi_{20} Z_0 + \pi_{21} Z_1 + \pi_{22} Z_2 + \pi_{23} Z_3 + \pi_{24} Z_4 + \pi_{25} Z_5 + v_2$$

and in matrix form:

$$Y_1 = Z \pi_1 + v_1$$

(13.6)

$$Y_2 = Z \pi_2 + v_2$$

where Z is the matrix of observations of all predetermined variables in the model.

π_1 is the vector of parameters of the first stage equation for Y_1.

π_2 is the vector of parameters of the first stage equation for Y_2.

v_1 is the vector of diatrubances of the first stage equation for Y_1.

v_2 is the vector of disturbances of the first stage equation for Y_2.

Note that the equations of (13.5) and (13.6) are of the form where one endogenous variable is a function of predetermined variables and a disturbance term. TSLS entails applying least squares methods in two stages. The first stage involves estimating the parameters of (13.6) by ordinary least squares and then obtaining the predicted values of the endogenous variables for observed values of the predetermined variables. In the second stage of estimation the observed values of the endogenous variables on the right side of (13.4) are replaced by their predicted values from the first stage and then least squares methods are applied to obtain estimates of the structural parameters.

The least squares estimators of the structural parameters of (13.6) are as follow:

$$\begin{aligned} \hat{\pi}_1 &= (Z'Z)^{-1} Z'Y_1 \\ \hat{\pi}_2 &= (Z'Z)^{-1} Z'Y_2 \end{aligned} \tag{13.7}$$

and the predicted values of the endogenous variables are:

$$\begin{aligned} \hat{Y}_1 &= Z\,\hat{\pi}_1 = Z(Z'Z)^{-1} Z'Y_1 \\ \hat{Y}_2 &= Z\,\hat{\pi}_2 = Z(Z'Z)^{-1} Z'Y_2 \end{aligned} \tag{13.8}$$

All of the statistical quantities related to least squares procedures and tests for testing hypotheses about regression parameters are applicable to the first stage euqations. This means quantities such as R^2 and $(1 - R^2)$ are valid and tests such as t and F may be used to test hypotheses about the π's.

For the second stage the predicted values of (13.8) are substituted into (13.4) to yield:

$$\begin{aligned} Y_1 &= \beta_{12} \hat{Y}_2 + \gamma_{11} Z_1 + \gamma_{12} Z_2 + \gamma_{13} Z_3 + \gamma_{10} Z_0 + u_1^* \\ Y_2 &= \beta_{21} \hat{Y}_1 + \gamma_{21} Z_1 + \gamma_{24} Z_4 + \gamma_{25} Z_5 + \gamma_{20} Z_0 + u_2^* \end{aligned} \tag{13.9}$$

The second stage involves applying ordinary least squares procedures to the equation of (13.9) to obtain estimates of the structural parameters. The estimators are biased but consistent. R^2 is no longer a strictly valid statistical quantity. Conventional t and F tests are not strictly valid for testing hypotheses about the structural parameters. Furthermore, the residuals variance and variances of coefficients obtained by estimating (13.9) by ordinary least squares are not appropriate after estimates of β's and γ's are obtained, $\hat{Y}_2$ and $\hat{Y}_1$ on the right side of (13.9) have to be replaced with observed values of Y_2 and Y_1 to obtain predicted values of Y_1 and Y_2 on the left-hand side of (13.9). Then (Y_1 - this predicted Y_1) = $\hat{u}_1$ and (Y_2 - this predicted Y_2) = $\hat{u}_2$ which then can be used to estimate the appropriate residual sums of squares and appropriate residual variances. Using these residual variances the appropriate variances and standard errors of the estimated coefficients may be obtained.

The General Case

Again let us consider the general model

(13.10) $Y \beta + Z \Gamma = U$

consisting of M structural equations. From this system of equations let us focus on the gth equation which is:

(13.11) $$\beta_{g1} Y_{1t} + \beta_{g2} Y_{2t} + - - - + \beta_{gm_1} Y_{m_1 t} + \gamma_{g1} Z_{1t} + - - - + \gamma_{gn_1} Z_{n_1} + \gamma_{g0} Z_0 = U_{gt}$$

where t = 1, 2, . . . T. T is the sample size. Normalizing (13.11) with respect to Y_{1t} by letting $\beta_{g1} = -1$ we have:

(13.12) $$Y_{1t} = \beta_{g2} Y_{2t} + - - - + \beta_{gm_1} Y_{m_1 t} + \gamma_{g1} Z_{1t} + - - - + \gamma_{gn_1} Z_{n_1} + \gamma_{g0} Z_0 - U_{gt}$$

In matrix form the normalized gth equation is:

$$(13.13) \quad Y_1 = Y_{g\Delta}\,\beta_{g\Delta} + Z_{g*}\,\gamma_{g*} - u_g$$

where:

$$Y_1 = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \cdot \\ \cdot \\ \cdot \\ Y_{t1} \end{bmatrix} \qquad Y_{g\Delta} = \begin{bmatrix} Y_{12} & Y_{13} - - - & Y_1{}^{m_1} \\ Y_{22} & Y_{23} - - - & Y_2{}^{m_1} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ Y_{T2} & Y_{T3} - - - & Y_T{}^{m_1} \end{bmatrix}$$

$$\beta_{g\Delta} = \begin{bmatrix} \beta_{g2} \\ \beta_{g3} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{gm_1} \end{bmatrix} \qquad Z_{g*} = \begin{bmatrix} Z_{11} - - - Z_{1n_1} & Z_{10} \\ Z_{21} - - - Z_{2n_1} & Z_{20} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ Z_{T1} - - - Z_{Tn_1} & Z_{T0} \end{bmatrix}$$

$$\gamma_{g*} = \begin{bmatrix} \gamma_{g1} \\ \gamma_{g2} \\ \cdot \\ \cdot \\ \gamma_{gn_1} \\ \gamma_{g0} \end{bmatrix} \qquad u_g = \begin{bmatrix} u_{1g} \\ u_{2g} \\ \cdot \\ \cdot \\ u_{Tg} \end{bmatrix}$$

$Y_{g\Delta}$ represents the matrix of observations of endogenous variables in the equation other than the one with respect to which the equation was normalized. $\beta_{g\Delta}$ is the vector of parameters of the $Y_{g\Delta}$ endogenous variables. Z_{g*} denotes

the matrix of observations of the predetermined variables in the gth equation. A "dummy variable" for the intercept is included in Z_{g*}. γ_{g*} is the vector of parameters of the predetermined variables in the gth equation. u_g is the vector of disturbances for the gth equation.

In (13.10) the matrix of predetermined variables is:

$$Z = [Z_{g*} \; Z_{g**}] \tag{13.14}$$

where Z_{g**} is a T x $(N - n_1)$ matrix of observations of predetermined variables excluded from the gth equation but in the remainder of the model.

The first stage equations are as follow:

$$\begin{aligned} Y_2 &= Z\pi_2 + v_2 \\ Y_3 &= Z\pi_3 + v_3 \\ &\vdots \\ Y_{m_1} &= Z\pi_{m_1} + v_{m_1} \end{aligned} \tag{13.15}$$

or

$$Y_{g\Delta} = Z\pi + v \tag{13.16}$$

where

$$\pi = \begin{bmatrix} \pi_2 \\ \pi_3 \\ \vdots \\ \pi_{m1} \end{bmatrix} \quad \text{and} \quad v = \begin{bmatrix} v_2 \\ v_3 \\ \vdots \\ v_{m1} \end{bmatrix}$$

Least squares estimators of the parameters of the first stage equations are:

$$
\begin{aligned}
\hat{\pi}_2 &= (Z'Z)^{-1} Z'Y_2 \\
\hat{\pi}_3 &= (Z'Z)^{-1} Z'Y_3 \\
&\vdots \\
\hat{\pi}_{m_1} &= (Z'Z)^{-1} Z'Y_{m_1}
\end{aligned}
\tag{13.17}
$$

or

$$\hat{\pi} = (Z'Z)^{-1} Z'Y_{g\Delta} \tag{13.18}$$

Predicted values or the endogenous variables are:

$$
\begin{aligned}
\hat{Y}_2 &= Z \hat{\pi}_2 \\
\hat{Y}_3 &= Z \hat{\pi}_3 \\
&\vdots \\
\hat{Y}_{m_1} &= Z \hat{\pi}_{m_1}
\end{aligned}
$$

or

$$\hat{Y}_{g\Delta} = Z \hat{\pi} \tag{13.20}$$

The predicted values may also be expressed in terms of the observed endogenous variables and the residuals of the first stage equations as follows:

$$\hat{Y}_{g\Delta} = Y_{g\Delta} - \hat{V} \tag{13.21}$$

where

$$\hat{V} = Y_{g\Delta} - \hat{Y}_{g\Delta} = Y_{g\Delta} - Z \hat{\pi}$$

Now substituting $\hat{Y}_{g\Delta}$ for $Y_{g\Delta}$ in (13.13) we have:

$$Y_1 = \hat{Y}_{g\Delta} \beta_{g\Delta} + Z_{g*} \gamma_{g*} - w_g \tag{13.22}$$

where w_g is the disturbance for the gth equation when the observed values of the endogenous variables on the right are replaced with their respective

predicted values. Hence,

$$(13.23)\quad Y_1 = \hat{Y}_{g\Delta}\, Z_{g^*} \begin{bmatrix} \beta_{g\Delta} \\ \gamma_{g^*} \end{bmatrix} - w_g$$

The least squares estimators of the structural parameters of the gth equation are:

$$(13.24)\quad \begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hat{\gamma}_{g^*} \end{bmatrix} = \{ \left[\hat{Y}_{g\Delta}\ \ Z_{g^*}\right]' \left[\hat{Y}_{g\Delta}\ Z_{g^*}\right] \}^{-1} \left[\hat{Y}_{g\Delta}\ Z_{g^*}\right]'\ Y_1$$

$$= \begin{bmatrix} \hat{Y}'_{g\Delta}\ \hat{Y}_{g\Delta} & \hat{Y}'_{g_\Delta}\ Z_{g^*} \\ Z'_{g^*}\ \hat{Y}_{g\Delta} & Z'_{g^*}\ Z_{g^*} \end{bmatrix}^{-1} \left[\hat{Y}_{g\Delta}\ Z_{g^*}\right]'\ Y_1$$

Substituting (13.21) into (13.24) yields:

$$(13.25)\quad \begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hat{\gamma}_{g^*} \end{bmatrix} = \begin{bmatrix} (Y_{g\Delta} - \hat{V})'\ (Y_{g\Delta} - \hat{V}) & (Y_{g\Delta} - \hat{V})\ Z_{g^*} \\ Z'_{g^*}\ (Y_{g\Delta} - \hat{V}) & Z'_{g^*}\ Z_{g^*} \end{bmatrix}^{-1} \begin{bmatrix} (Y_{g\Delta} - \hat{V})' \\ Z'_{g^*} \end{bmatrix} Y_1$$

Now

$$(13.26)\quad (Y_{g\Delta} - \hat{V})'(Y_{g\Delta} - \hat{V}) = Y'_{g\Delta}\, Y_{g\Delta} - Y'_{g\Delta}\ \hat{V} - \hat{V}'Y_{g\Delta} + \hat{V}'\hat{V}$$

$$= Y'_{g\Delta}\, Y_{g\Delta} - \hat{V}'\hat{V}$$

because $\hat{Y}_{g\Delta}$ and $\hat{V}$ are uncorrelated since $\hat{Y}_{g\Delta} = Z\,\hat{\pi}$ and Z is uncorrelated with $\hat{V}$. Furthermore,

$$(13.27)\quad Z'\hat{V} = Z'(Y_{g\Delta} - Z(Z'Z)^{-1}\, Z'Y_{g\Delta})$$

$$= Z'Y_{g\Delta} - Z'\, Z(Z'Z^{-1})\, Z'Y_{g\Delta} = Z'\, Y_{g\Delta} - Z'\, Y_{g\Delta}$$

$$= 0$$

Hence,

$$(13.28)\quad Z'_{g^*}\, \hat{V} = 0$$

since it is a submatrix of $Z'\hat{V}$ and thus must be a submatrix of the null matrix. Therefore, the TSLS estimators of the parameters of the gth structural equation are:

$$(13.29)\quad \begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hat{\gamma}_{g*} \end{bmatrix} = \begin{bmatrix} Y'_{g\Delta} Y_{g\Delta} - \hat{V}'\hat{V} & Y'_{g\Delta} Z_{g*} \\ Z'_{g*} Y_{g\Delta} & Z'_{g*} Z_{g*} \end{bmatrix}^{-1} \begin{bmatrix} (Y_{g\Delta} - V)' \\ Z'_{g*} \end{bmatrix} Y_1$$

The residual for the gth structural equation is:

$$(13.30)\quad \hat{u}_g = (Y_1 - \hat{Y}_1) = Y_1 - \begin{bmatrix} Y_{g\Delta} Z_{g*} \end{bmatrix} \begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hat{\gamma}_{g*} \end{bmatrix}$$

and

$$(13.31)\quad \hat{u}'_g \hat{u}_g = (Y_1 - \hat{Y}_1)'(Y_1 - \hat{Y}_1) = \left\{ Y_1 - \begin{bmatrix} Y_{g\Delta} Z_{g*} \end{bmatrix} \begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hat{\gamma}_{g*} \end{bmatrix} \right\} \left\{ Y_1 - \begin{bmatrix} Y_{g\Delta} Z_{g*} \end{bmatrix} \begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hat{\gamma}_{g*} \end{bmatrix} \right\}$$

which is the sum of squared residuals from the gth structural equation based on the observed values of the endogenous variables on the right of the equality sign. The residual variance is

$$(13.32)\quad s_g^2 = \frac{\hat{u}'_g \hat{u}_g}{T - (m_1 + n_1)}$$

where there are $(m_1 - 1)$ endogenous variables but $(n_1 + 1)$ predetermined variables including the "dummy variable" for the intercept. Following least squares procedures the estimated variance-covariance matrix of the estimators of the structural parameters of the gth equation is:

$$(13.33)\quad \text{Var} \begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hat{\gamma}_{g*} \end{bmatrix} = s_\zeta^2 \begin{bmatrix} Y'_{g\Delta} Y_{g\Delta} - \hat{V}'\hat{V} & Y'_{g\Delta} Z_{g*} \\ Z'_{g*} Y_{g\Delta} & Z'_{g*} Z_{g*} \end{bmatrix}^{-1}$$

Testing and Interpretation

Although frequently the t-tests, F-tests, and R^2 are presented for the second stage equations of TSLS they are not strictly valid for the estimated structural equations. Tests for serial correlation aren't strictly legitimate. All of these should be interpreted with caution. Remember we are dealing with biased estimators and the estimated variances and standard errors of the estimated coefficients are estimates of asymptotic variances and standard errors, not estimates of variances and standard errors for finite sample sizes. They may serve as useful approximations and should be recognized as approximations. Often the magnitude of the estimated coefficient relative to its estimated standard error is used to judge whether or not a parameter is different from zero. If the absolute value of the estimated coefficient is greater than its estimated standard error the parameter is considered different from zero.

Care should be exercised in interpreting the estimated coefficients of structural equations particularly the coefficients of endogenous variables. There is some difficulty with the partial regression coefficient interpretation of these estimates of parameters. This is not unique to TSLS but is true for any estimators of structural parameters of a simultaneous equation system. For the coefficients of endogenous variables, the condition that everything else is invariant is untenable. A change in any endogenous variable leads to simultaneous changes in other endogenous variables. Other endogenous variables cannot be kept invariant when one is permitted to vary. This doesn't mean we don't attempt to interpret these coefficients but simply suggests that we should recognize their limitations. For very small or infinitesimal changes this is probably not a serious matter. For large change as we often consider for policy purpose this may be quite critical. The same holds for elasticities or other economic quantities which might be computed or derived from the estimates of structural parameters.

CHAPTER 14

Limited Information Single Equation Estimation

Introduction

The limited information single equation (LISE) method for estimating parameters of a simultaneous system of linear equations was developed prior to the two-stage least squares method. It is somewhat of a more difficult estimation procedure to understand, it is somewhat more complex and the computational task is considerably more laborious. However, computer programs for estimating parameters by LISE are available so that computational consideration in choosing a method of estimation are of less importance.

Two criteria of estimation may be employed in estimating parameters of a simultaneous system of linear equations LISE and both criteria result in the same estimating formulas. The two criteria are the maximum likelihood and least variance ratio. A more detailed discussion of the maximum likelihood principle will be presented in a later chapter although some use of the criterion will be made here. We will draw some on the least variance ratio (LVR) but recognize that in terms of the formulas used in estimation it doesn't make any difference which criterion is used.

Least-Variance Ratio Criterion

Consider a simultaneous model of linear equations such as:

(14.1) $Y\beta + Z\Gamma = U$

where:

Y is a (T x M) matrix of observations of endogenous variables.
β is a (M x M) matrix of parameters of the endogenous variables.
Z is a (T x N) matrix of observations of predetermined variables.
Γ is a (N x M) matrix of the predetermined variables.
U is a (T x M) matrix of disturbances.

The model is complete in the sense that there are M endogenous variables and M equations.

The gth equation of this system is:

$$(14.2)\quad \beta_{g1} Y_{1t} + \beta_{g2} Y_{2t} + \cdots + \beta_{gm_1} Y_{m_1 t} + \gamma_{g1} Z_{1t} + \cdots + \gamma_{gn_1} Z_{n_1} + \gamma_{g0} Z_0 = u_{gt}$$

where $t = 1, 2, \ldots, T$ is the sample size. In matrix form the gth equation is:

$$(14.3)\quad Y_{g\Delta}\, \beta_{g\Delta} + Z_{g*}\, \gamma_{g*} = u_g$$

where:

$$Y_{g\Delta} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m_1} \\ Y_{21} & Y_{22} & \cdots & Y_{2m_1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ Y_{T_1} & Y_{T_2} & \cdots & Y_{Tm_1} \end{bmatrix} \qquad \beta_{g\Delta} = \begin{bmatrix} \beta_{g1} \\ \beta_{g2} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{gm_1} \end{bmatrix}$$

$$Z_{g*} = \begin{bmatrix} Z_{11} & \cdots & Z_{1n_1} & Z_{10} \\ Z_{21} & \cdots & Z_{2n_1} & Z_{20} \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ Z_{T_1} & \cdots & Z_{Tn_1} & Z_{T0} \end{bmatrix}$$

$$\gamma_{g*} = \begin{bmatrix} \gamma_{g1} \\ \gamma_{g2} \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{gn_1} \\ \gamma_{g0} \end{bmatrix} \qquad u_g = \begin{bmatrix} u_{1g} \\ u_{2g} \\ \cdot \\ \cdot \\ \cdot \\ u_{Tg} \end{bmatrix}$$

$Y_{g\Delta}$ represents the matrix of observations of endogenous variables in the equation. At this point no distinction is made between the one with respect to which the equation is normalized and the other endogenous variables in the gth equation. $\beta_{g\Delta}$ is the vector of parameters of the endogenous variables. No normalization is indicated but normalization simply means that one of the parameters of $\beta_{g\Delta}$ will assume a value of either 1 or -1. Z_{g*} denotes the matrix of observations of the predetermined varisbles in the gth equation including a dummy variable for the intercept. Y_{g*} is the vector of the parameters of the n_1 predetermined variables in the gth equation. u_g is the vector of disturbances for the gth equation.

In (14.1) the matrix of observations of predetermined variables in the model is:

$$(14.4) \qquad Z = \left[Z_* \; Z_{**}\right]$$

where

Z_* is the T x n_1 matrix of observations of the predetermined variables in the gth structural equation.

Z_{**} is the T x (N - n_1) matrix of observations of predetermined variables included in the model but not appearing in the gth equation.

Now for convenience purposes let us define the linear combination of endogenous variables in the gth equation as:

$$(14.5) \qquad \tilde{Y} = \begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{Y}_T \end{bmatrix} = Y_{g\Delta}\,\beta_{g\Delta}$$

The least variance ratio criterion involves estimating the coefficients of the endogenous variables in the gth equation, $\beta_{g\Delta}$, so that the residual variance when $\tilde{Y} = Y_{g\Delta}\,\beta_{g\Delta}$ is regressed on Z_* relative to the residual variance when

$\widetilde{Y} = Y_{g\Delta}\,\beta_{g\Delta}$ is regressed on $Z = [Z_* \; Z_{**}]$ is made as small as possible. In other words the addition of the predetermined variables excluded from the gth equation but included in the remainder of the model, Z_{**}, should make a minimal improvement in the explained sum of squares of $\widetilde{Y}$.

The least-variance ratio criterion is compatible with the restrictions on parameters implied by the structural equation with respect to the reduced form. (14.2) could be rewritten as follows:

$$(14.6) \quad \widetilde{Y} + Y_{g**}\,[0] + Z_*\,\gamma_* + Z_{**}\,[0] = u_g$$

where

Y_{g**} is the matrix of observations of endogenous variables in the system other than those appearing in the gth equation.

$[0]$ are null matrices of appropriate order.

Thus the corresponding reduced form implied by these restrictions on the parameters are:

$$(14.7) \quad \widetilde{Y} = Z_*\,\gamma_* + Z_{**}\,[0] + w_g$$

where w_g is the reduced form disturbance. Thus the predetermined variables $Z_{n_1+1}, Z_{n_1+2}, - - -, Z_N$ do not appear in the gth equation and the implication is that if we were to compute the regression of $\widetilde{Y}$ on all Z's in the model including those that do not appear in the gth equation then we should expect to get zeroes for the estimated parameters of Z_{**}, Z_{n_1+1}, Z_{n_1+2}, - - -, Z_N. But because of sampling variation, it is impossible to choose values of $\beta_{g\Delta}$ that will make all estimated regression coefficients of Z_{**} be exactly zero for a finite sample in an over-identified equation. The next best thing is to choose the values of $\beta_{g\Delta}$ such that the excluded predetermined variables are as unimportant as possible which is exactly what the least-variance ratio criterion attempts to do. In other words the values of $\beta_{g\Delta}$ are chosen in such a manner that the contribution of Z_{**} to accounting for the sum of squares $\widetilde{Y}$ above the amount already accounted for by Z_* is a minimum.

Assumptions

The assumptions employed in estimation by the limited information single equation method are the same as those used in TSLS. The relevant assumptions are given in (13.2) of the preceeding chapter, and will not be repeated here.

Estimation Procedure

The variance ratio which is to minimized is obtained by regressing the composite endogenous variable, $\tilde{Y}$, on Z_* and on Z and forming the ratio of the first residual sum of squares over the second. The sum of squares of the composite endogenous variable is:

$$(14.8) \quad \tilde{Y}'\tilde{Y} = \beta_{g\Delta}' \; Y_{g\Delta}' \; Y_{g\Delta} \; \beta_{g\Delta}$$

Now if we regress Y on Z_* we are concerned with the following regression equation:

$$(14.9) \quad \tilde{Y} = Z_* \pi_{\Delta *} + v_*$$

where v_* is a vector of serially independent disturbances and $E(v_*) = 0$. Ordinary least squares yields the following estimators of the parameters of (14.9):

$$(14.10) \quad \hat{\pi}_{\Delta *} = (Z_*' Z_*)^{-1} Z_*' \tilde{Y}$$

The explained sum of squares is:

$$(14.11) \quad SSR_* = \hat{\pi}_{\Delta *}' \; Z_*' \; \tilde{Y} - n\bar{\tilde{Y}}^2 = \tilde{Y}' \; Z_* \; (Z_*' \; Z_*)^{-1} \; Z_*' \tilde{Y} - n\bar{\tilde{Y}}^2$$

where $\bar{\tilde{Y}} = \sum_{i-1}^{n} \frac{Y_i}{n}$

The residual sum of squares is:

$$(14.12) \quad SSE_* = Y'Y - n\bar{\tilde{Y}}^2 - SSR_*$$

$$= \tilde{Y}'\tilde{Y} - n\bar{\tilde{Y}}^2 - \tilde{Y}' \; Z_*(Z_*' \; Z_*)^{-1} \; Z_*' \; \bar{\tilde{Y}} + n\bar{\tilde{Y}}^2$$

$$= \tilde{Y}'\tilde{Y} - \tilde{Y}'Z_* \; (Z_*' \; Z_*)^{-1} \; Z_*' \; \tilde{Y}$$

$$= \beta_{g\Delta}' \; Y_{g\Delta}' \; Y_{g\Delta} \; \beta_{g\Delta} - (Y_{g\Delta} \; \beta_{g\Delta})'Z_*(Z_*'Z_*)^{-1}Z_*' \; (Y_{g\Delta} \; \beta_{g\Delta})$$

$$= \beta'_{g\Delta} \; Y'_{g\Delta} \; Y_{g\Delta} \beta_{g\Delta} - \beta'_{g\Delta} \; Y'_{g\Delta} \; Z_* (Z'_* \; Z_*)^{-1} \; Z'_* \; Y_{g\Delta} \beta_{g\Delta}$$

$$SSE_* = \beta'_{g\Delta} \; W^*_{\Delta\Delta} \; \beta_{g\Delta}$$

where

$$(14.13) \quad W^*_{\Delta\Delta} = Y'_{g\Delta} \; Y_{g\Delta} - Y'_{g\Delta} \; Z_* (Z'_* \; Z_*)^{-1} \; Z'_* \; Y_{g\Delta}$$

and $W^*_{\Delta\Delta}$ is $m_1 \times m_1$.

Similarily we may regress $\widetilde{Y}$ on $Z = [Z_* Z_{**}]$. The regression model is:

$$(14.14) \quad \widetilde{Y} = Z \pi_{\Delta **} + v_{**}$$

where v_{**} is a vector of serially independent disturbances and $E(v_{**}) = 0$. Ordinary least squares yields the following estimators of the parameters of (14.14):

$$(14.15) \quad \hat{\pi}_{\Delta **} = (Z'Z)^{-1} \; Z \; \widetilde{Y}$$

The explained sum of squares is:

$$(14.16) \quad SSR_{**} = \hat{\pi}_{\Delta **} Z'\widetilde{Y} - n \, \bar{\bar{Y}}^2 = \widetilde{Y}'Z \; (Z'Z)^{-1} \; Z'\widetilde{Y} - n \, \bar{\bar{Y}}^2$$

where $\bar{\bar{Y}} = \sum_{i=1}^{n} \frac{\widetilde{Y}_i}{n}$

The residual sum of squares is:

$$(14.17) \quad SSE_{**} = \widetilde{Y}'\widetilde{Y} - n \, \bar{\bar{Y}}^2 - SSR^{**}$$

$$= \widetilde{Y}'\widetilde{Y} - n \, \bar{\bar{Y}}^2 - \widetilde{Y}'Z(Z'Z)^{-1} \; Z' \; \widetilde{Y} + n \, \bar{\bar{Y}}^2$$

$$= \widetilde{Y}'\widetilde{Y} - \widetilde{Y}'Z(Z'Z)^{-1} \; Z'\widetilde{Y}$$

$$= \beta'_{g\Delta} \; Y'_{g\Delta} \; Y_{g\Delta} \beta_{g\Delta} - \beta'_{g\Delta} \; Y'_{g\Delta} \; Z(Z'Z)^{-1} \; Z'Y_{g\Delta} \beta_{g\Delta}$$

$$= \beta'_{g\Delta} \; W_{\Delta\Delta} \beta_{g\Delta}$$

where

$$(14.18) \quad W_{\Delta\Delta} = Y'_{g\Delta} \; Y_{g\Delta} - Y'_{g\Delta} \; Z(Z'Z)^{-1} \; Z'Y_{g\Delta}$$

and $W_{\Delta\Delta}$ is $m_1 \times m_1$.

The least-variance ratio boils down to a choice of $\beta_{g\Delta}$ which minimizes:[1/]

$$(14.19) \quad \ell = \frac{\beta_{g\Delta}' \; W^*_{\Delta\Delta} \; \beta_{g\Delta}}{\beta_{g\Delta}' \; W_{\Delta\Delta} \; \beta_{g\Delta}}$$

Now the task is to minimize (14.19). This entails differentiating ℓ partially with respect to the individual parameters of $\beta_{g\Delta}$, i.e., β_{gj} $(j = 1,2,\ldots,m_1)$ and equating these first order partial derivatives to zero. This yields the following m_1 equations which may be solved simultaneously:

$$(14.20) \quad w^*_j \beta_{g\Delta} - \ell \, w_j \beta_{g\Delta} = 0$$

where w^*_j denotes the jth row of $W^*_{\Delta\Delta}$.

w_j denotes the jth row of $W_{\Delta\Delta}$.

Expressing these equations in matrix form yields:

$$(14.21) \quad (W^*_{\Delta\Delta} - \ell \; W_{\Delta\Delta}) \; \beta_{g\Delta} = 0$$

These may be obtained directly by matrix differentiation:

$$(14.22) \quad \frac{\partial \ell}{\partial \beta_{g\Delta}} = \frac{2W^*_{\Delta\Delta} \beta_{g\Delta}(\beta_{g\Delta}' \; W_{\Delta\Delta} \; \beta_{g\Delta}) - (\beta_{g\Delta}' \; W^*_{\Delta\Delta} \; \beta_{g\Delta}) \; (2) \; W_{\Delta\Delta} \; \beta_{g\Delta}}{(\beta_{g\Delta}' \; W_{\Delta\Delta} \; \beta_{g\Delta})^2}$$

and

$$(14.23) \quad W^*_{\Delta\Delta} \; \beta_{g\Delta} \; (\beta_{g\Delta}' \; W_{\Delta\Delta} \; \beta_{g\Delta}) - (\beta_{g\Delta}' \; W^*_{\Delta\Delta} \; \beta_{g\Delta}) \; W_{\Delta\Delta} \; \beta_{g\Delta} = 0$$

[1/] The set of estimators of parameters of endogenous variables in the gth equation which minimize the variance ratio also fulfill the maximum likelihood conditions. The $\beta_{g\Delta}$ which minimize (14.19) also maximize

$L = -\frac{1}{2} \log_e \frac{\beta_{g\Delta}' \; W^*_{\Delta\Delta} \; \beta_{g\Delta}}{\beta_{g\Delta}' \; W_{\Delta\Delta} \; \beta_{g\Delta}}$ which is the objective function that the single equation maximum likelihood procedure reduces to. See J. Johnston, Econometric Methods, 1972, pp. 384-387.

thus

$$(14.24)\quad \left(W^*_{\Delta\Delta} - \frac{\beta'_{g\Delta}\, W^*_{\Delta\Delta}\, \beta_{g\Delta}}{\beta'_{g\Delta}\, W_{\Delta\Delta}\, \beta_{g\Delta}}\, W_{\Delta\Delta}\right) \beta_{g\Delta} = 0$$

hence

$$(14.25)\quad (W^*_{\Delta\Delta} - \ell\, W_{\Delta\Delta})\, \beta_{g\Delta} = 0$$

Now (14.21) and (14.25) have a non-trivial solution for $\beta_{g\Delta}$ only if the determinental equation

$$(14.26)\quad |W^*_{\Delta\Delta} - \ell\, W_{\Delta\Delta}| = 0$$

is satisfied. The elements of both $W^*_{\Delta\Delta}$ and $W_{\Delta\Delta}$ are functions of sample observations as given in (14.13) and (14.18). Hence (14.26) provides us with a polynomial in ℓ (m_1 order) which must be solved for the smallest root, $\hat{\ell}$. There are as many non-zero roots, $\hat{\ell}$, as the rank of $\left[W^*_{\Delta\Delta} - \ell\, W_{\Delta\Delta}\right]$ which in this case is m_1 the number of endogenous variables in the gth equation.

After the smallest root has been obtained it is substituted into (14.25) to obtain

$$(14.27)\quad (W^*_{\Delta\Delta} - \hat{\ell}\, W_{\Delta\Delta})\, \hat{\beta}_{g\Delta} = 0$$

This is a system of m_1 equation in m_1 unknown, the $\hat{\beta}_{g\Delta}$, thus we can solve for the estimated coefficients of the endogenous variables in the gth equation. But since the system of equations given in (14.27) is equal to a null vector and $\hat{\beta}_{g\Delta} \neq 0$, the β_{gj}'s are in proportional terms to each other. To obtain a unique set of estimates of $\beta_{g\Delta}$ a restriction in the form of "fixing" the value of one of the parameters has to be utilized. Normalization provides this restriction. If one of the β_{gj}'s is set equal to -1, then $\hat{\beta}_{g\Delta}$ is unique.

Now we are in a position to obtain the estimates of the parameters of the predetermined variables in the gth equation. From the specifications of the gth equation in (14.3) we may solve for:

$$(14.28)\quad Z_* \gamma_* = - Y_{g\Delta}\, \beta_{g\Delta} + u_g$$

and

$$(14.29)\quad Z_\star' Z_\star \gamma_\star = - Z_\star' Y_{g\Delta} \beta_{g\Delta} + Z_\star' u_g$$

and

$$(14.30)\quad \hat{\gamma}_\star = -(Z_\star' Z_\star)^{-1} Z_\star' Y_{g\Delta} \hat{\beta}_{g\Delta} = - \hat{\pi}_{\Delta\star}$$

where $\hat{\pi}_{\Delta\star}$ may be obtained from (14.10) after $\hat{\beta}_{g\Delta}$ is known. Also, note that (14.30) may be written:

$$(14.31)\quad \hat{\gamma}_\star = - \left[(Z_\star' Z_\star)^{-1} Z_\star' Y_{g\Delta}\right] \hat{\beta}_{g\Delta}$$

where $\left[(Z_\star' Z_\star)^{-1} Z_\star' Y_{g\Delta}\right]$ are least squares estimators obtained when the endogenous variables in the gth equation are regressed on the predetermined variables in the gth equation, $Z_\star$. This regression model is:

$$(14.32)\quad Y_{g\Delta} = Z_\star \pi^\star + w_g$$

and

$$(14.33)\quad \hat{\pi}_\star = (Z_\star' Z_\star)^{-1} Z_\star' Y_{g\Delta}$$

where $\hat{\pi}_\star$ is a $n_1 \times m_1$ matrix consisting of m_1 column vectors each consisting of n_1 elements. In other words the equations corresponding to (14.33) are:

$$(14.34)\quad \hat{\pi}_{1\star} = (Z_\star' Z_\star)^{-1} Z_\star Y_{g1}$$

$$\hat{\pi}_{2\star} = (Z_\star' Z_\star)^{-1} Z_\star' Y_{g2}$$

$$\vdots \qquad\qquad \vdots$$

$$\hat{\pi}_{m_1\star} = (Z_\star' Z_\star)^{-1} Z_\star' Y_{gm_1}$$

and

$$(14.35)\quad \hat{\pi}_\star = \left[\hat{\pi}_{1\star}\, \hat{\pi}_{2\star}, \quad , \hat{\pi}_{m_1\star}\right]$$

Thus (14.31) is also,

$$(14.36)\quad \hat{\gamma}_\star = - \hat{\pi}_\star \hat{\beta}_{g\Delta}$$

An estimate of the residual variance may be obtained from:

$$(14.37) \quad \hat{\sigma}^2_g = \frac{\hat{u}'_g \quad \hat{u}_g}{T - (m_1 + n_1)}$$

where $\hat{u}_g = Y_{g\Delta} \hat{\beta}_{g\Delta} + Z_* \hat{\gamma}_*$

The asymptotic variances and covariances of the limited information single equation estimators are

$$(14.38) \quad \text{Var}\,(\hat{\beta}_{g\Delta}\ ,\ \hat{\gamma}_*) = \frac{\hat{\ell} - 1}{T - (m_1 + n_1)} \quad \beta'_{g\Delta} \quad W_{\Delta\Delta} \quad \beta_{g\Delta} \quad \begin{bmatrix} F_{\beta\beta} & -F_{\beta\gamma} \\ -F_{\beta\gamma} & F_{\gamma\gamma} \end{bmatrix}$$

where $W_{\Delta\Delta} = Y'_{g\Delta}\ Y_{g\Delta} - Y_{g\Delta}\ Z(Z'Z)^{-1}\ Z'Y_{g\Delta}$

$\hat{\ell}$ is the smallest root of $|W^*_{\Delta\Delta} - \ell W_{\Delta\Delta}| = 0$

$$F_{\beta\beta} = \left[{}_{gg}\{W^*_{\Delta\Delta} - W_{\Delta\Delta}\} - \frac{\ell - 1}{\hat{\beta}'_{g\Delta}\ W_{\Delta\Delta}\ \hat{\beta}_{g\Delta}} \quad {}_{gg}\{(\hat{\beta}_{g\Delta} W_{\Delta\Delta})' \quad \hat{\beta}_{g\Delta}\ W_{\Delta\Delta}\} \right]^{-1}$$

${}_{gg}\{W^*_{\Delta\Delta} - W_{\Delta\Delta}\}$ denotes that the gth row and gth column have been deleted. Here g refers to the endogenous variable with respect to which the equation was normalyzed. If it was Y, then g = 1 and gg = 1,1.

Now,

$$(14.39) \quad \text{Var}\,(\hat{\beta}_{g\Delta}) = \frac{(\hat{\ell}-1)\ \hat{\beta}'_{g\Delta}\ W_{\Delta\Delta}\ \hat{\beta}_{g\Delta}}{T - (m_1 + n_1)} \quad F_{\beta\beta}$$

are the estimated variances and covariances of $\hat{\beta}_{g\Delta}$.

The estimated covariances between $\hat{\beta}_{g\Delta}$ and $\hat{\gamma}_*$ are:

$$(14.40) \quad \text{cov}\,(\hat{\beta}_{g\Delta}, \hat{\gamma}_*) = \frac{(\hat{\ell} - 1)\ \hat{\beta}_{g\Delta}\ W_{\Delta\Delta}\ \hat{\beta}_{g\Delta}}{T - (m_1 + n_1)} \quad (-F_{\beta\gamma})$$

where $F_{\beta\gamma} = F_{\beta\beta go} \left[(Z_*'Z_*)^{-1} Z_*' Y_{g\Delta}\right]'$

$go\left[(Z_*'Z_*)^{-1} Z_*' Y_{g\Delta}\right]'$ denotes that the gth row has been deleted.

The estimated variances and covariances of $\hat{\gamma}_*$ are:

$$(14.41)\quad \text{var}(\hat{\gamma}_*) = \frac{(\hat{\ell} - 1)\,\hat{\beta}_{g\Delta}\, W_{\Delta\Delta}\, \hat{\beta}_{g\Delta}}{T-(m_1+n_1)}\, F_{\gamma\gamma}$$

where

$$F_{\gamma\gamma} = (Z_*' Z_*)^{-1} + {}_{og}\left[(Z_*' Z_*)^{-1} Z_*' Y_{g\Delta}\right] F_{\beta\gamma}$$

$$= (Z_*' Z_*)^{-1} + {}_{og}\left[(Z_*'Z_*)^{-1} Z_*' Y_{g\Delta}\right] F_{\beta\beta}\; go\left[(Z_*'Z_*)^{-1} Z_*'Y_{g\Delta}\right]'$$

and ${}_{og}\left[(Z_*' Z_*)^{-1} Z_*' Y_{g\Delta}\right]$ denotes the gth column has been deleted and

$go\left[(Z_*'Z_*)^{-1} Z_*' Y_{g\Delta}\right]'$ denotes that the gth row of the transposed matrix has been deleted.

CHAPTER 15

k Class, Double k and h Class Estimators

Introduction

Theil developed k-class and h-class estimators which are general classes of limited information estimators.[1/] Nagar[2/] a student of Theil's and later a colleague of his developed the double k-class estimators of which the k-class are a special case. All three of these classes of estimators are general classes of limited information with TSLS and LISE estimators being special cases. Each of these methods is briefly described below. The purpose of the presentation is to make you cognizent of these estimators and to make you aware of the relationships between the limit information estimators.

k Class Estimators

Again consider the general model

(15.1) $Y \beta + Z \Gamma = U$

consisting of M structural equations. Let us focus on the gth equation which when normalized with respect to Y_1 is:

(15.2) $Y_1 = Y_{g\Delta} \beta_{g\Delta} + Z_{g*} Y_{g*} - u_g$

1/ Henri Theil, Economic Forecasts and Policy (Contributions to Economic Analysis No. 15) Amsterdam, North-Holland Publishing COmpany, first edition, 1958; second edition, 1961, pp. 231-233 for k-class and pp. 353-355 for h-class.

2/ A. L. Nagar, "Double k-Class Estimators of Parameters in Simultaneous Equations and Their Small Sample Properties, International Economic Review, Vol. 3, No. 2 (May) 1962, pp. 168-188.

where: Y_1 is the vector of observations of the endogenous variable with respect to which the gth equation is normalized.

$Y_{g\Delta}$ is the matrix of observations of endogenous variables in the gth equation beside the one with respect to which the equation is normalized.

$\beta_{g\Delta}$ is the vector of parameters of $Y_{g\Delta}$.

Z_{g*} is the matrix of observation of the predetermined variables in the gth equation and includes a variable for the intercept.

γ_{g*} is the vector of parameters the predetermined variables in the gth equation.

u_g is the vector of disturbances of the gth equation.

Assume the sample size is T and there are adequate degrees of freedom for estimation. Also,consider (15.1) to be over-identified. The matrix of observation of the total number of predetermined variables in (15.1), M, is:

$$(15.3) \quad Z = \left[Z_{g*} \; Z_{g**} \right]$$

where Z_{g**} is T x $(N - n_1)$ represents the observations of the $(N - n_1)$ predetermined variables excluded from the gth equation.

In developing the two stage least square estimators (13.15)-(13.28) we reached one formulation given in (13.29) which is repeated here as:

$$(15.4) \quad \begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hat{Y}_{g*} \end{bmatrix} = \begin{bmatrix} Y'_{g\Delta} \; Y_{g\Delta} - \hat{V}'\hat{V} & Y'_{g\Delta} \; Z_{g*} \\ Z'_{g*} \; Y_{g\Delta} & Z'_{g*} \; Z_{g*} \end{bmatrix} -1 \begin{bmatrix} (Y_{g\Delta} - \hat{V})' \\ Z'_{g*} \end{bmatrix} Y_1$$

where $\hat{V}$ is the matrix of residuals of the first stage equations of the TSLS procedure or in more general terms may be viewed as the matrix of residuals of the reduced forms. The essence of k-class estimators may be seen from the following expression of the k-class estimators:

$$(15.5)\qquad \begin{bmatrix} \hat{\beta}^{(k)}_{g\Delta} \\ \hat{\gamma}^{(k)}_{g^*} \end{bmatrix} = \begin{bmatrix} Y'_{g\Delta}Y_{g\Delta} - k\,\hat{V}'\hat{V} & Y'_{g\Delta}Z_{g^*} \\ Z'_{g^*}Y_{g\Delta} & Z'_{g^*}Z_{g^*} \end{bmatrix}^{-1} \begin{bmatrix} (Y_{g\Delta} - k\,\hat{V})' \\ Z'_{g^*} \end{bmatrix} Y_1$$

where k is a / variable. The different types of k-class estimators are defined by k taking on different values.

k-class estimators are consistent if the probability limit of k is equal to 1. Note that ordinary least squares, two stage least squares and limited information single equation (least-variance ratio or maximum likelihood) estimators are particular k-class estimators. For ordinary least squares k=0 and the least squares estimators are not consistent. For TSLS, k=1 and the estimators are consistent because plim k=1. Now for LISE estimators $k = \hat{\ell}$ where $\hat{\ell}$ is the smallest root of the determinantal equation (14.26) presented here as:

$$(15.6)\qquad |\, W^*_{\Delta\Delta} - \ell\, W_{\Delta\Delta} \,| = 0$$

Remember that the smallest root, $\hat{\ell}$, of (15.6) minimizes the least variance ratio and also maximizes the appropriate maximum likelihood function. The probability limit of $\hat{\ell}$ has been shown to be 1 and hence the LISE estimators are consistent.

Possibly a simplier way of viewing LISE estimators is through the k-class framework. After one has solved (15.6) for $\hat{\ell}$ then the LISE estimators may be obtained by letting $k = \ell$ in (15.5) and solve for the estimators. Hence,

$$(15.7)\qquad \begin{bmatrix} \hat{\beta}^{(\hat{\ell})}_{g\Delta} \\ \hat{\gamma}^{(\hat{\ell})}_{g^*} \end{bmatrix} = \begin{bmatrix} Y'_{g\Delta}Y_{g\Delta} - \hat{\ell}\,\hat{V}'\hat{V} & Y'_{g\Delta}Z_{g^*} \\ Z'_{g^*}Y_{g\Delta} & Z'_{g^*}Z_{g^*} \end{bmatrix}^{-1} \begin{bmatrix} (Y_{g\Delta} - \hat{\ell}\,\hat{V})' \\ Z'_{g^*} \end{bmatrix} Y_1$$

Now the residual for the gth structural equation is:

$$(15.8)\quad \hat{u}_g = Y_1 - \left[Y_{g\Delta}\ Z_{g*}\right] \begin{bmatrix} \hat{\beta}_{g*}^{(\ell)} \\ \hat{\gamma}_{g\Delta}^{(\ell)} \end{bmatrix}$$

and the residual sum of squares is:

$$(15.9)\quad \hat{u}_g'\ \hat{u}_g = \left\{Y_1 - \left[Y_{g\Delta}Z_{g*}\right] \begin{bmatrix} \hat{\beta}_{g\Delta}^{(\ell)} \\ \hat{\gamma}_{g\Delta}^{(\ell)} \end{bmatrix}\right\}' \left\{Y_1 - \left[Y_{g\Delta}Z_{g*}\right] \begin{bmatrix} \hat{\beta}_{g\Delta}^{(\ell)} \\ \hat{\gamma}_{g\Delta}^{(\ell)} \end{bmatrix}\right\}$$

The residual variance for the gth equation is obtained by dividing (15.9) by the appropriate degrees of freedom which is $\left[T - (m_1 + n_1)\right]$ and

$$(15.10)\quad s_g^2 = \frac{\hat{u}_g'\ \hat{u}_g}{T - (m_1 + n_1)}$$

The variance-covariance matrix of the LISE estimators, as with any of the k-class estimators, may be obtained from:

$$(15.11)\quad \text{Var} \begin{bmatrix} \hat{\beta}_{g\Delta}^{(\ell)} \\ \hat{\gamma}_{g*}^{(\ell)} \end{bmatrix} = \underset{T\to\infty}{\text{plim}} \left\{ \left(\begin{bmatrix} \hat{\beta}_{g\Delta}^{(\ell)} \\ \hat{\gamma}_{g*}^{(\ell)} \end{bmatrix} - \begin{bmatrix} \beta_{g\Delta}^{(\ell)} \\ \gamma_{g*}^{(\ell)} \end{bmatrix} \right) \left(\begin{bmatrix} \hat{\beta}_{g\Delta}^{(\ell)} \\ \hat{\gamma}_{g*}^{(\ell)} \end{bmatrix} - \begin{bmatrix} \beta_{g\Delta}^{(\ell)} \\ \gamma_{g*}^{(\ell)} \end{bmatrix} \right)' \right\}$$

which yields

$$(15.12)\quad \text{Var} \begin{bmatrix} \hat{\beta}_{g\Delta}^{(\ell)} \\ \hat{\gamma}_{g*}^{(\ell)} \end{bmatrix} = s_g^2 \begin{bmatrix} Y_{g\Delta}' Y_{g\Delta} - \hat{\ell}\,\hat{V}'\hat{V} & Y_{g\Delta}' Z_{g*} \\ Z_{g*}' Y_{g\Delta} & Z_{g*}' Z_{g*} \end{bmatrix}^{-1}$$

Theil's k-class estimators define a family of estimators of which ordinary least squares, two stage least sqaures, and limited information single equation (least-variance ratio or maximum likelihood) are special cases. Other criteria of estimation may result in other values of k and hence other types of k-class estimators. Hence, it is a very general framework which defines a large family of estimators for simultaneous systems of equations.

However, regardless of the values of k its probability limit has to be 1 if the estimators are to be consistent.

h-Class Estimator

The h-class estimators define another alternative set of estimation methods for parameter of simultaneous equation systems. This method was also developed by Theil. The estimators are based on an arbitrary real number h. The general h-class estimators for (15.2) are defined as follows:

$$(15.13)\quad \begin{bmatrix} \hat{\beta}_g(h) \\ \hat{\gamma}_{g*}(h) \end{bmatrix} = \begin{bmatrix} Y'_{g\Delta}Y_{g\Delta}-(1-h^2)\,\hat{V}'\hat{V} & Y'_{g\Delta}\,Z_{g*} \\ Z'_{g*}\,Y_{g\Delta} & Z'_{g*}\,Z_{g*} \end{bmatrix} -1 \begin{bmatrix} (Y_{g\Delta}-(1-h)V)' \\ Z'_{g*} \end{bmatrix} y_1$$

Both ordinary least squares and two stage least squares are special cases of the h-class estimaters. For ordinary least squares h=1. For two stage least squares h=0. It follows from the consideration of k-class estimators that if the h-class estimators are to be consistent then the probability limit of (1-h) has to be 1. Again this is not true for ordinary least squares but plim $(1-h^2)$ = plim (1-h) = 1 for two stage least squres and the estimators are consistent.

Note that limited information single equation estimator (LVR or ML) are not one of the h-class estimators. h has to be a real number and $\hat{\lambda}$ is the smallest root of a determinantal equation and is a real number. But for the LISE (LVR or ML) to be a special case of the h-class estimators $(1-h^2)$ = (1-h) would have to hold and this is untenable.

The residual variance of the gth equation and the variance-covariance matrix of the h-class estimators may be obtained in a similar manner as was done for two stage least squares (13.31), (13.32) and (13.33) and k-class LISE, (15.9), (15.10) and (15.12).

Double k-Class Estimators

Nagar introduced the double k-class estimators which is an even more general class of estimators for parameters of systems of simultaneous linear equations. The double k-class estimators encompass both the k-class and the h-class estimators. The crux of the double k-class estimators is that instead of a single parameter, k, as in the k-class estimators we may consider two k parameters, k_1 and k_2. k_1 and k_2 are two arbitrary real numbers which may be equal to each other or different from each other.

The double k-class estimators for the gth equation, (15.2) are defined as:

$$(15.14)\quad \begin{bmatrix} \hat{\beta}_{g\Delta}(k_1, k_2) \\ \hat{\gamma}_{g*}(k_1, k_2) \end{bmatrix} = \begin{bmatrix} Y'_{g\Delta} Y_{g\Delta} - k_1 \hat{V}'\hat{V} & Y'_{g\Delta} Z_{g*} \\ Z'_{g*} Y_{g\Delta} & Z'_{g*} Z_{g*} \end{bmatrix}^{-1} \begin{bmatrix} (Y_{g\Delta} - k_2\hat{V})' \\ Z'_{g*} \end{bmatrix} Y_1$$

The double k-class estimators are consistent if plim k_1 = plim k_2 = 1. Note that if $k_1 = (1-h^2)$ and $k_2 = (1-h)$ that the double k-class estimators are the h-class estimators. If $k_1 = k_2 = k$ the double k-class estimators are the k-class estimators. If $k_1 = k_2 = 0$ the method yields ordinary least squares estimators which are inconsistent. If $k_1 = k_2 = 1$ the two stage least squares estimators are obtained. If $k_1 = k_2 = \hat{\ell}$ the limited information single equation estimators are obtained. These latter two estimators are consistent.

The procedure for obtaining the residual variance and variance-covariance matrix of the estimators when $k_1 = k_2 = k$ has already been discussed. The asymptotic variance-covariance matrix and some small sample properties of the double k-class estimators have been worked out by Nagar.[3/]

3/ Nagar, Ibid and A. L. Nagar, "The Bias and Moment Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations", Econometrica, Vol. 27, No. 4 (October), pp. 575-595.

CHAPTER 16

Instrumental Variables

Introduction

The instrumental variable method of estimating parameters of a simultaneous system of linear equations is also a limited information method of estimation. It is applicable to a single equation of a system of equations. It was one of the first methods used in estimating parameters of a simultaneous systems of linear equations which utilized the over-identifying constraint.

The method involve what its name implies. Predetermined variables are used as instruments. When focusing on a given structural equation one chooses predetermined variables excluded from that equation but included in the other equations of the model as instruments. The choice of instrumental variables is arbitrary. One instrumental variable is chosen for each endogenous variable in the equation under consideration except the endogenous variable with respect to which the equation is normalized. The equation is multiplied through by each instrument and then sumed over the sample observations. Alternatively, if the multiplication is performed in terms of vectos the sum over sample observations is encompassed in the vector products. Also, the equation under consideration is multiplied through by the predetermined variables in that equation. The result is a set of linear equations similar to the normal euqations of the least squares procedure. There will be as many equations as there are unknown parameters and the set of equations may be solved out for the unknown parameters.

Choice of Instruments

Although the choice of instruments is arbitrary but bias will be minimized if the instruments possess certain attributes. First, the instrumental variables should be truly exogenous. Thus lagged endogenous variables should no be used as instruments. Second, the instrumental variables should

show considerable variation. If there is a choice between two or more instruments the one exhibiting the greatest variation should be chosen. Third, the instrumental variables should be measured without error. If possible do not choose exogenous variables which you suspicion may contain measurement error. Fourth, the endogenous variables should react more strongly to each other then to the instrumental variables. Finally, the set of instrumental variables should be as free from inter-correlation as possible. Avoid choosing a highly correlated set of instrumental variables.

One of the difficulties with the instrumental variable method is that it is difficult to fulfill these conditions. Furthermore, it may be difficult to verify whether or not the instrumental variables possess the desired attributes.

One of the biggest objections to the instrumental variable method is that the solution for the unknown parameters is not unique unless the system of simultaneous linear equations is just-identified. In the case of an over-identified equation there is more than one set of instruments from which to choose. The different sets of instrumental variables will yield different estimators of the parameters. All the estimators will be biased but all will be consistent estimators.

Just-Identified System

Let us consider a two equation system of linear equations which is just-identified such as the following:

$$(16.1) \quad Y_1 = Y_2 \beta_{12} + Z_1 \gamma_{11} + Z_2 \gamma_{12} + Z_0 \gamma_{10} + u_1$$

$$Y_2 = Y_1 \beta_{21} + Z_1 \gamma_{21} + Z_3 \gamma_{23} + Z_0 \gamma_{20} + u_1$$

Suppose we are interested in estimating the parameters of the first equation. Let us assume that all Z's are exogenous. Our choice of instruments in this case is from the Z's excluded from the first equation but included in the second. Z_3 is the only variable which can be used as an instrument. Using

Z_3 as an instrument we form the following set of equations:

$$
\begin{aligned}
Z_3'Y_1 &= Z_3'Y_2\beta_{12} + Z_3'Z_1\,\gamma_{11} + Z_3'Z_2\,\gamma_{12} + Z_3'Z_0\,\gamma_{10} + Z_3'\,u_1 \\
Z_1'Y_1 &= Z_1'Y_2\beta_{12} + Z_1'Z_1\,\gamma_{11} + Z_1'Z_2\,\gamma_{12} + Z_1'Z_0\,\gamma_{10} + Z_1'\,u_1 \\
Z_2'Y_1 &= Z_2'Y_2\beta_{12} + Z_2'Z_1\,\gamma_{11} + Z_2'Z_2\,\gamma_{12} + Z_2'Z_0\,\gamma_{10} + Z_2'\,u_1 \\
Z_0'Y_1 &= Z_0'Y_2\beta_{12} + Z_0'Z_1\,\gamma_{11} + Z_0'Z_1\,\gamma_{12} + Z_0'Z_0\,\gamma_{10} + Z_0'\,u_1
\end{aligned}
\tag{16.2}
$$

Rewriting (16.2) in matrix form we have:

$$
\begin{bmatrix} Z_3'Y_1 \\ Z_1'Y_1 \\ Z_2'Y_1 \\ Z_0'Y_1 \end{bmatrix} = \begin{bmatrix} Z_3'Y_2 & Z_3'Z_1 & Z_3'Z_2 & Z_3'Z_0 \\ Z_1'Y_2 & Z_1'Z_1 & Z_1'Z_2 & Z_1'Z_0 \\ Z_2'Y_2 & Z_2'Z_1 & Z_2'Z_2 & Z_2'Z_0 \\ Z_0'Y_2 & Z_0'Z_1 & Z_0'Z_2 & Z_0'Z_0 \end{bmatrix} \begin{bmatrix} \beta_{12} \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} Z_3'\,u_1 \\ Z_1'\,u_1 \\ Z_2'\,u_1 \\ Z_0'\,u_1 \end{bmatrix}
\tag{16.3}
$$

and the instrumental variable estimators are:

$$
\begin{bmatrix} \hat{\beta}_{12} \\ \hat{\gamma}_{11} \\ \hat{\gamma}_{12} \\ \hat{\gamma}_{10} \end{bmatrix} = \begin{bmatrix} Z_3'Y_2 & Z_3'Z_1 & Z_3'Z_2 & Z_3'Z_0 \\ Z_1'Y_2 & Z_1'Z_1 & Z_1'Z_2 & Z_1'Z_0 \\ Z_2'Y_2 & Z_2'Z_1 & Z_2'Z_2 & Z_2'Z_0 \\ Z_0'Y_2 & Z_0'Z_1 & Z_0'Z_2 & Z_0'Z_0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} Z_3'Y_1 \\ Z_1'Y_1 \\ Z_2'Y_1 \\ Z_0'Y_1 \end{bmatrix}
\tag{16.4}
$$

In a similar manner the estimators for the second equation of (16.1) is:

$$
\begin{bmatrix} \hat{\beta}_{21} \\ \hat{\gamma}_{21} \\ \hat{\gamma}_{23} \\ \hat{\gamma}_{20} \end{bmatrix} = \begin{bmatrix} Z_2'Y_1 & Z_2'Z_1 & Z_2'Z_3 & Z_2'Z_0 \\ Z_1'Y_1 & Z_1'Z_1 & Z_1'Z_3 & Z_1'Z_0 \\ Z_3'Y_1 & Z_3'Z_1 & Z_3'Z_3 & Z_3'Z_0 \\ Z_0'Y_1 & Z_0'Z_1 & Z_0'Z_3 & Z_0'Z_0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} Z_2'Y_2 \\ Z_1'Y_2 \\ Z_3'Y_2 \\ Z_0'Y_2 \end{bmatrix}
\tag{16.5}
$$

Note how similar (16.4) and (16.5) are to ordinary least squares estimators. For a just-identified system such as (16.1) the instrumental variable estimators are identical to indirect least squares, two stage least squares ,

limited information LVR or ML, k-class, couble k-class, h-class, three stage least squares, and full information maximum likelihood.

Over-Identified Systems

Now let us consider an over-identified simultaneous system of linear equations in which there are M endogenous variables and equation and N predetermined variables. Let us assume all N predetermined variables are exogenous. The model is:

(16.6) $Y\beta = Z\Gamma + U$

Let us focus on estimating the parameters of the gth equation using the instrumental variable technique. The gth structural equation normalized on Y_1 is:

(16.7) $Y_1 = Y_{g\Delta}\,\beta_{g\Delta} + Z_{g*}\,\gamma_{g*} + u_g$

Let us assume that Z_* contains a dummy variable to allow for the estimation of an intercept. Now,

(16.8) $Z = \left[Z_{g*}\ Z_{g**}\right]$

where Z_{g**} are the predetermined variables excluded from the gth equation but included in the other equations of the model.

The instruments will be chosen from Z_{g**}. Assuming there are m_1 endogenous variables in the gth equation we will need to choose (m_1-1) instruments from Z_{g**}. Let us denote the chosen instruments as

(16.9) $Z'_{g\Delta}$

Now form the equations:

(16.10) $Z'_{g\Delta}\,Y_1 = Z'_{g\Delta}Y_{g\Delta}\beta_{g\Delta} + Z'_{g\Delta}Z_{g*}\gamma_{g*} + Z'_{g\Delta}\,u_g$

$Z'_{g*}\,Y_1 = Z'_{g*}Y_{g\Delta}\beta_{g\Delta} + Z'_{g*}Z_{g*}\gamma_{g*} + Z'_{g*}\,u_g$

and

$$(16.11)\qquad \begin{bmatrix} Z'_{g\Delta} \\ \hline Z'_{g*} \end{bmatrix} Y_1 = \begin{bmatrix} Z'_{g\Delta} Y_{g\Delta} & Z'_{g\Delta} Z_{g*} \\ \hline Z'_{g*} Y_{g\Delta} & Z'_{g*} Z_{g*} \end{bmatrix} \begin{bmatrix} \beta_{g\Delta} \\ \hline \gamma_{g*} \end{bmatrix} + \begin{bmatrix} Z'_{g\Delta} u_g \\ \hline Z'_{g*} u_g \end{bmatrix}$$

The instrumental variable estimators of the parameters of the gth equation are:

$$(16.12)\qquad \begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hline \hat{\gamma}_{g*} \end{bmatrix} = \begin{bmatrix} Z'_{g\Delta} Y_{g\Delta} & Z'_{g\Delta} Z_{g*} \\ \hline Z'_{g*} Y_{g\Delta} & Z'_{g*} Z_{g*} \end{bmatrix}^{-1} \begin{bmatrix} Z'_{g\Delta} \\ \hline Z'_{g*} \end{bmatrix} Y_1$$

The instrumental variable estimators of the other structural equations may be obtained in a similar manner.

Concluding Comment

One advantage of the instrumental variable method is that the computations are only slightly more burdensome than for ordinary least squares. Furthermore, it is relatively easy to understand the notion of instrumental variables. However, the instrumental variable method has no advantages over the other methods presented earlier. In addition, in an over-identified system the instrumental variable estimators are not unique with respect to alternative sets, of instruments. Thus their greatest usefulness may be in obtaining "quick" results in exploratory analyses.

Three-Stage Least Squares

Introduction

Three-stage least squares is a full information method for estimating structural parameters of an over-identified system of simultaneous linear equations. The method was developed by Zellner and Theil and is an extension of the two-stage least squares method.[1/] In the first stage of two-stage least squares the moment matrix of the reduced-form disturbances is estimated and in the second stage the parameters of a single structural equation are estimated after an adjustment has been made by the estimated moment matrix of the reduced form disturbances. This adjustment is given in (13.25) and the second stage estimators are given in (13.29). The mechanics of two-stage least squares estimation basically entail replacing the endogenous variables of a structural equation, other than the one with respect to which the equation is normalized, with predicted values from the first stage and applying ordinary least squares again to obtain estimates of the structural parameters of that equation. The three-stage least squares estimation goes one step further and uses the estimated variance-covariance matrix of structural disturbances to estimate all coefficients of the entire system simultaneously.

The three stage least squares method has full information if the variance-covariance matrix of the structural disturbances is not diagonal (structural disturbances have non-zero "contemporaneous" covariances) in that the estimation of coefficients of any identifiable equation gains in efficiency as soon as there are other equations that are over identified. For a just-identified system or for any structural equation for which all other structural equations are just-identified there is no gain in efficiency. The three-

---

1/ Arnold Zellner and H. Theil, "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations", Econometrica, Vol. 30, No. 1, January 1962, pp. 54-78.

stage method is general and allows for restrictions of parameters in different structural equations. The computational procedure is rather simple except that a rather sizeable matrix has to be inverted.

Estimation Procedure

Let us again consider the over-identified system of simultaneous linear equations:

$$Y\,\beta + Z\,\Gamma = U \tag{17.1}$$

consisting of M equations with M endogenous variables and N predetermined variables including a dummy variable for the intercept. Let us assume a sample of T of adeuqate size to perform all necessary estimation. The gth equation normalized with respect to Y, is:

$$Y_1 = Y_{g\Delta}\,\beta_{g\Delta} + Z_{g*}\,\gamma_{g*} - u_g \tag{17.2}$$

where $Y_{g\Delta}$ includes $(m_1 - 1)$ endogenous variables and Z_{g*} consists of n_1 predetermined variables including a discrete variable for an intercept, Z_{g**} includes $(N - n_1)$ predetermined variables. Hence,

$$Z = \left[Z_{g*}\ Z_{g**} \right] \tag{17.3}$$

The equation used in the second stage of estimation of two-stage least squares is:

$$Y_1 = \hat{Y}_{g\Delta}\,\beta_{g\Delta} + Z_{g*}\,\gamma_{g*} - w_g \tag{17.4}$$

where w_g is the disturbance of the gth structural equation when $\hat{Y}_{g\Delta}$ has been substituted for $Y_{g\Delta}$. The two-stage least squares estimators of the parameters of the gth equation are:

$$\begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hat{\gamma}_{g*} \end{bmatrix} = \begin{bmatrix} \hat{Y}'_{g\Delta}\hat{Y}_{g\Delta} & \hat{Y}'_{g\Delta}Z_{g*} \\ Z'_{g*}\hat{Y}_{g\Delta} & Z'_{g*}Z_{g*} \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_{g\Delta} & Z_{g*} \end{bmatrix}' Y_1 \tag{17.5}$$

where $\hat{Y}_{g\Delta}$ are predicted values of the endogenous variables obtained from the first stage equations. The predicted values are obtained from

$$(17.6) \quad \hat{Y}_{g\Delta} = Z\,\hat{\pi}$$

where

$$(17.7) \quad \hat{\pi} = (Z'Z)^{-1} Z'\,Y_{g\Delta}$$

hence,

$$(17.8) \quad \hat{Y}_{g\Delta} = Z(Z'Z)^{-1} Z'Y_{g\Delta}$$

and substituting into (17.4) we have:

$$(17.9) \quad \begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hat{\gamma}_{g*} \end{bmatrix} = \begin{bmatrix} [Z(Z'Z)^{-1} Z'Y_{g\Delta}]' [Z(Z'Z)^{-1} Z'Y_{g\Delta}] & [Z(Z'Z)^{-1} Z'Y_{g\Delta}]' Z_{g*} \\ Z'_{g*} [Z(Z'Z)^{-1} Z'Y_{g\Delta}] & Z'_{g*} Z_{g*} \end{bmatrix}^{-1}$$

$$\times \quad \{[Z(Z'Z)^{-1} Z'Y_{g\Delta}] Z_{g*}\}' Y_1$$

where the estimators are expressed in terms of the observed values of the predetermined variables in the entire system including the gth equation and the endogenous variables in the gth equation. The estimators for each equation in the system may be obtained by repeating the process for each equation. The variance-covariance matrix of the estimators of the gth equation is:

$$(17.10) \quad \text{Var}\begin{bmatrix} \hat{\beta}_{g\Delta} \\ \hat{\gamma}_{g*} \end{bmatrix} = s_{gg} \begin{bmatrix} [Z(Z'Z)^{-1}Z'Y_{g\Delta}]' [Z(Z'Z)^{-1}Z'Y_{g\Delta}] & [Z(Z'Z)^{-1}Z'Y_{g}]'Z_{g*} \\ Z'g_{*} [Z(Z'Z)^{-1} Z'Y_{g\Delta}] & Z'_{g*} Z_{g*} \end{bmatrix}^{-1}$$

where s_{gg} is the residual variance of the gth equation and is:

$$(17.11) \quad s_{gg} = \frac{\hat{u}'_g \quad \hat{u}_g}{T-(m_1 + n_1)}$$

where

$$(17.12)\quad \hat{u}_g'\hat{u}_g = \{Y_1 - [Y_{g\Delta}Z_{g^*}]\begin{bmatrix}\hat{\beta}_{g\Delta}\\ \hat{\gamma}_{g^*}\end{bmatrix}\}' \{Y_1 - [Y_{g\Delta}Z_{g^*}]\begin{bmatrix}\hat{\beta}_{g\Delta}\\ \hat{\gamma}_{g^*}\end{bmatrix}\}.$$

This procedure may be applied to other equations in the system to obtain the estimators, the variance-covariance matrix of the estimators, residual variances of the other equations, and estimates of the covariances between the residuals of the various equations. Consider the gth and hth equations, $g \neq h$, the covariance between their residuals may be estimated as follows:

$$(17.13)\quad s_{gh} = \frac{\hat{u}_g'\hat{u}_h}{\sqrt{[T-(m_1+n_1)]_g[T-(m_1+n_1)]_h}} = \frac{\hat{u}_h'\hat{u}_g}{\sqrt{[T-(m_1+n_1)]_h[T-(m_1+n_1)]_g}}$$

where $T - [(m_1 + n_1)]_g$ is the degrees of freedom of the gth equation and $T - [(m_1 + n_1)]_h$ is the degrees of freedom of the hth equation. Alternatively, T, may be used as the degrees of freedom and the choice between these two alternatives depends on their small sample properties about which little is known except in the case where all explanatory variables are exogenous. That is the case of ordinary least squares and if T is used as the denominator of (17.11) and (17.13) the resulting estimators are biased. In that case the denominators given in (17.11) and (17.13) are appropriate.

Note that in terms of the observations and estimates of structural parameters the covariance between the gth and hth equation is:

$$(17.14)\quad s_{gh} = \{Y_g - [Y_{g\Delta}Z_{g^*}]\begin{bmatrix}\hat{\beta}_{g\Delta}\\ \hat{\gamma}_{g^*}\end{bmatrix}\}' \{Y_h - [Y_{h\Delta}Z_{h^*}]\begin{bmatrix}\hat{\beta}_{h\Delta}\\ \hat{\gamma}_{h^*}\end{bmatrix}\}$$

$$\div \sqrt{[T - (m_1 + n_1)]_g [T - (m_1 + n_1)]_h}$$

The basic notion underlying three-stage least squares is to apply generalized least squares to the whole set of structural equations to estimate all parameters simultaneously. Beyond the two-stage least squares part of the procedure this entails re-estimating the parameters of the structural equations of the system taking into acount, in a generalized least squares framework, the estimated variances and covariances of the residuals of the structural equations which are generated in the second stage of estimation. To show this we need to relabel our matrices and vectors of estimators in order to achieve a more compact form.

Let us now consider the first structural equation of (17.1) as:

$$(17.15)\quad Y_1 = Y_{1\Delta}\,\beta_{1\Delta} + Z_{1*}\,\gamma_{1*} - u_1 = \begin{bmatrix} Y_{1\Delta} & Z_{1*} \end{bmatrix} \begin{bmatrix} \beta_{1\Delta} \\ \gamma_{1*} \end{bmatrix} - u_1$$

and let

$$(17.16)\quad X_1 = \begin{bmatrix} Y_{1\Delta} & Z_{1*} \end{bmatrix} \quad \text{and} \quad \delta = \begin{bmatrix} \beta_{1\Delta} \\ \gamma_{1*} \end{bmatrix}$$

hence

$$(17.17)\quad Y_1 = X_1\,\delta_1 - u_1$$

All M structural equations may be relabelled in this form. Now if we premultiply (17.17) by Z' where $Z = \begin{bmatrix} Z_{1*} & Z_{1**} \end{bmatrix}$ we have:

$$(17.18)\quad Z'Y_1 = Z'X_1\delta_1 - Z'u_1$$

The variance-covariance matrix of $Z'u_1$ is:

$$(17.19)\quad E\left[(Z'u_1)\,(Z'u_1)'\right] = E\left[Z'u_1\,u_1'\,Z\right] = \sigma_{11}\,Z'Z$$

where σ_{11} is the variance of the first structural equation. Empirically one would estimate σ_{11} by the procedure given in (17.11).

Two-stage least squares estimators are obtained by applying least squares to (17.18) which yields

$$(17.20)\quad \hat{\delta}_1 = \left[X_1' Z(Z'Z)^{-1} Z' X_1\right]^{-1} \left[X_1' Z(Z'Z)^{-1} Z' Y_1\right]$$

and the estimated variance-covariance matrix of the estimators is:

$$(17.21)\quad \text{Var}(\hat{\delta}_1) = E\left[(\hat{\delta}_1 - \delta_1)(\hat{\delta}_1 - \delta_1)'\right] = s_{11}\left[X_1' Z(Z'Z)^{-1} Z'X\right]^{-1}$$

where s_{11} is the estimate of σ_{11} and is obtained from (17.11). Three-stage least squares estimators are obtained by applying GLS to (17.20).

Now we may write the entire system of structural equations in the form of (17.18) as:

$$(17.22)\quad \begin{bmatrix} Z' Y_1 \\ Z' Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Z' Y_M \end{bmatrix} = \begin{bmatrix} Z'X_1 & 0 & 0 & \cdots & 0 \\ 0 & Z'X_2 & 0 & \cdots & 0 \\ 0 & 0 & Z'X_3 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & Z'X_M \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \cdot \\ \cdot \\ \cdot \\ \delta_M \end{bmatrix} - \begin{bmatrix} Z' u_1 \\ Z' u_2 \\ \cdot \\ \cdot \\ \cdot \\ Z' u_M \end{bmatrix}$$

which is a system of NM equations involving

$$(17.23)\quad P = \sum_{i=1}^{M} (m_i + n_i)$$

parameters.

Now rewriting (17.22) in general matrix form we have:

$$(17.24)\quad Z'Y = \left[Z' X\right] \delta - Z' U$$

where:

$$Z'Y = \begin{bmatrix} Z'Y_1 \\ Z'Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Z'Y_M \end{bmatrix} \qquad Z'X = \begin{bmatrix} Z'X_1 & 0 & 0 & \cdots & 0 \\ 0 & Z'X_2 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & Z'X_M \end{bmatrix}$$

$$\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{bmatrix} \qquad Z'U = \begin{bmatrix} Z' u_1 \\ Z' u_2 \\ \vdots \\ Z' u_M \end{bmatrix}$$

Generalized least squares may be applied to (17.22) to estimate all elements of δ simultaneously. For this purpose the variance-covariance matrix of the disturbance vector in (17.22) is needed. This variance-covariance matrix is:

$$(17.25) \quad \text{Var} \begin{bmatrix} Z' u_1 \\ Z' u_2 \\ \vdots \\ Z' u_M \end{bmatrix} = \begin{bmatrix} \sigma_{11}Z'Z & \sigma_{12}Z'Z & \cdots & \sigma_{1M}Z'Z \\ \sigma_{21}Z'Z & \sigma_{22}Z'Z & \cdots & \sigma_{2M}Z'Z \\ \vdots & & & \\ \sigma_{M1}Z'Z & \sigma_{M2}Z'Z & \cdots & \sigma_{MM}Z'Z \end{bmatrix}$$

and

$$(17.26) \quad \text{Var} \begin{bmatrix} Z' u_1 \\ Z' u_2 \\ \vdots \\ Z' u_M \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & & & \\ \sigma_{M1} & \sigma_{M2} & & \sigma_{MM} \end{bmatrix} \otimes Z'Z$$

which is a Kronecker product. The σ_{gh}'s, $g \neq h$ are the covariances of the structural disturbances of the gth and hth equations. When $g = h$, σ_{gh} is the variance of the structural equation. The inverse of (17.25) is also needed and is:

$$(17.27)\qquad \text{Var}^{-1}\begin{bmatrix} Z'u_1 \\ Z'u_2 \\ \cdot \\ \cdot \\ \cdot \\ Z'u_M \end{bmatrix} = \begin{bmatrix} \sigma^{11}(Z'Z)^{-1} & \sigma^{12}(Z'Z)^{-1} & \cdots & \sigma^{1M}(Z'Z)^{-1} \\ \sigma^{21}(Z'Z)^{-1} & \sigma^{22}(Z'Z)^{-1} & \cdots & \sigma^{2M}(Z'Z)^{-1} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \sigma^{M1}(Z'Z)^{-1} & \sigma^{M2}(Z'Z)^{-1} & \cdots & \sigma^{MM}(Z'Z)^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^{11} & \sigma^{12} & & \sigma^{1M} \\ \sigma^{21} & \sigma^{22} & & \sigma^{2M} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \sigma^{M1} & \sigma^{M2} & & \sigma^{MM} \end{bmatrix} \otimes (Z'Z)^{-1}$$

where

$$(17.28)\qquad \begin{bmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1M} \\ \sigma^{21} & \sigma^{22} & \cdots & \sigma^{2M} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \sigma^{M1} & \sigma^{M2} & \cdots & \sigma^{MM} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{bmatrix}^{-1}$$

The σ^{gh}'s are elements of the inverse of the variance-covariance matrix of the disturbances.

Now the σ_{gh}'s are generally unknown but can be replaced by the estimated variances and covariances given by (17.11) and (17.13) which yields,

$$(17.29)\qquad \hat{\text{Var}}\begin{bmatrix} Z'u_1 \\ Z'u_2 \\ \cdot \\ \cdot \\ \cdot \\ Z'u_M \end{bmatrix} = \begin{bmatrix} s_{11}Z'Z & s_{12}Z'Z & \cdots & s_{1M}Z'Z \\ s_{21}Z'Z & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ s_{M1}Z'Z & & \cdots & s_{MM}Z'Z \end{bmatrix}$$

$$= \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1M} \\ s_{21} & s_{22} & \cdots & s_{2M} \\ \vdots & & & \vdots \\ s_{M1} & s_{M2} & \cdots & s_{MM} \end{bmatrix} \otimes (Z'Z)$$

and

$$(17.30)\quad \hat{\text{Var}}^{-1}\begin{bmatrix} Z'u_1 \\ Z'u_2 \\ \cdot \\ \cdot \\ \cdot \\ Z'u_M \end{bmatrix} = \begin{bmatrix} s^{11}(Z'Z)^{-1} & s^{12}(Z'Z)^{-1} & \cdots & s^{1M}(Z'Z)^{-1} \\ s^{21}(Z'Z)^{-1} & & \cdots & s^{2M}(Z'Z)^{-1} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ s^{M1}(Z'Z)^{-1} & & \cdots & s^{MM}(Z'Z)^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} s^{11} & s^{12} & \cdots & s^{1M} \\ s^{21} & s^{22} & \cdots & s^{2M} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ s^{M1} & s^{M2} & \cdots & s^{MM} \end{bmatrix} \otimes (Z'Z)^{-1}$$

Now applying the generalized least squares method utilizing (17.30) yields the third stage estimators:

$$(17.31)\quad \hat{\delta} = \begin{bmatrix} s^{11}X_1'Z(Z'Z)^{-1}Z'X_1 & s^{12}X_1'Z(Z'Z)^{-1}Z'X_2 & \cdots & s^{1M}X_1'Z(Z'Z)^{-1}Z'X_M \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ s^{M1}X_M'Z(Z'Z)^{-1}Z'X_1 & & \cdots & s^{MM}X_M'Z(Z'Z)^{-1}Z'X_M \end{bmatrix}^{-1}$$

$$X \begin{bmatrix} \Sigma s^{1g} X_1' & Z(Z'Z)^{-1} Z' Y_g \\ \\ \Sigma s^{Mg} X_M' & Z(Z'Z)^{-1} Z' Y_g \end{bmatrix}$$

where the summation is extended over $g = 1,2,\ldots,M$.

The variance-covariance matrix of $\hat{\delta}$ is:

$$(17.32) \quad \text{Var } (\hat{\delta}) = \begin{bmatrix} s^{11} X_1' Z(Z'Z)^{-1} Z' X_1 & - - - - & s^{1M} X_1' Z(Z'Z)^{-1} Z' X_M \\ \\ s^{M1} X_M' Z(Z'Z)^{-1} Z' X_1 & - - - - & s^{MM} X_M' Z(Z'Z)^{-1} Z' X_M \end{bmatrix}^{-1}$$

The gains in efficiency by using three stage least squares compared with two-stage least squares will occur only if the variance-covariance matrix of the residuals among the equations is not diagonal i.e., $[\sigma_{gh}]$ is not diagonal. If it is diagonal, two and three-stage least squares yield identical estimators.

CHAPTER 18

Maximum Likelihood Estimation

Introduction

Maximum likelihood method of estimation has widespread applications and the maximum likelihood criterion is one of the most widely used principles of estimation. Under certain assumptions, least squares estimators are a special case of maximum likelihood estimators. Using other criteria of estimation may yield maximum likelihood estimators and often maximum likelihood is used with reference to properties of estimators. The latter you will frequently see used in the literature when estimators obtained on the basis of a different criterion have the same desirable properties as the maximum likelihood estimators.

Maximum likelihood estimators have the desirable properties of being best, efficient, unbiased, consistent, and sufficient if some key assumptions are fulfilled. When maximum likelihood estimators are not unbiased often they can be readily modified and unbiased estimators obtained from the maximum likelihood estimators. Sometimes when maximum likelihood estimators are biased so are estimators of the same parameters that are based on other criteria. Also, even when maximum likelihood estimators are biased they may be asymptotically unbiased or consistent.

Maximum likelihood estimation is applicable to both single equations models and simultaneous systems of equations. Of course, it is much easier to apply this criterion to a single equation than to a set of simultaneous equations. But this is also true of other estimating criteria. In the following sections the logic of maximum likelihood estimation will be developed and an attempt will be made to show how maximum likelihood estimators are obtained and what assumptions are needed.

Likelihood and Probability

In concept likelihood is similar to probability but different. And that difference is an important thing in terms of understanding likelihood and maximum estimation.

Probability is a property of a sample. Given a probability density function with known (or estimated) values of the parameters, the probability or chance that a certain set of sample values will occur can be determined. Note the key is that given information about a probability density function the chances of occurrence or probability of a set of random variables may be determined. The point is for the selected functional form of the probability density function the parameters are known.

Likelihood is a property of the unknown parameter values. A sample of random variables is at hand but the parameters of the probability density which may have generated the particular sample are unknown. We are trying to determine the appropriate set of values of the parameters. We assume that to produce the sample, the data generating mechanism acted in the most likely fashion. Estimates of the parameters are chosen such that they are the most likely set to have been responsible for the particular sample of observations we have.

Likelihood Function

The likelihood function of n random variables $X_1\ X_2 - - - X_n$ (one X vsriable and n observations) is the joint density of the n random variables:

(18.1) $\quad g\ (X_1, X_2 - - - X_n;\ \theta)$

which is considered a function of θ. If $X_1\ X_2 - - - X_n$ is a random sample from the density $f(X;\theta)$ then the likelihood function is:

(18.2) $\quad g\ (X_1, X_2, - - - Z_n;\ \theta) = g_1\ (X_1;\theta) \cdot g_2(X_2;\theta) - - - \cdot g_n(X_n;\theta)$

The likelihood function (18.1) gives the relative likelihood that the random variables assume particular values of $X_1 - - - X_n$.

Suppose θ is unknown but we observe a particular sample $(X^*_1\ X^*_2 - - - X^*_n)$. We want to known from what density this particular set of values is most likely to have come. θ can take on many values but we are interested in θ which maximizes the likelihood of occurrence of the particular sample.

To obtain the maximum likelihood estimator, let us consider the following likelihood function:

$$(18.3) \quad L(\theta) = g(X_1\ X_2 - - - X_n;\ \theta)$$

If θ is the particular value of the parameters θ from all possible values of $\theta \epsilon \Omega$ which maximizes $L(\theta)$, then $\hat{\theta}$ is the maximum likelihood estimator of θ. In the case of a random sample from some density $f(X;\theta)$ the likelihood function may be separated into the n independent component functions and expressed as a product of the n independent components:

$$(18.4) \quad L(\theta) = f_1(X_1;\theta) \cdot f_2(X_s;\theta) - - - \cdot f_n(X_n;\theta)$$

The solution of the equation,

$$(18.5) \quad \frac{dL(\theta)}{d\theta} = 0$$

gives the maximum likelihood estimator of θ.[1/]

Maximum likelihood estimators possess an invariant property with respect to a logarithmic transformation (and some other types of transformations and operations) and $L(\theta)$ and $\log L(\theta)$ have their maxima at the same value of θ. It is often easier to find the maximum of the logarithmic likelihood.

If the likelihood function contains k parameters then

$$(18.6) \quad L(\theta_1, \theta_2 - - - \theta_k) = \prod_{i=1}^{n} f_i(X_i;\theta_1, \theta_2, - - - \theta_k)$$

[1/] Here and throughout we will assume that the likelihood functions satisfy the regularity conditions and in fact do have a maxima which may be found in this manner.

where $\prod_{i-1}^{n}$ represents a product of the n component functions. The maximum likelihood estimators of $\theta_1, \theta_2, - - - \theta_k$ are obtained by solving the following k equations:

$$\frac{\partial L(\theta_1, \theta_2, - - - \theta_k)}{\partial \theta_1} = 0$$

$$\frac{\partial L(\theta_1, \theta_2, - - - \theta_k)}{\partial \theta_2} = 0 \qquad (18.7)$$

$$\vdots$$

$$\frac{\partial L(\theta_1, \theta_2, - - - \theta_k)}{\partial \theta_k} = 0$$

In this case, it may be easier to work with the logarithm of the likelihood.

Now let us turn to some specific estimators such as the coefficients of a linear equation with which we might be confronted in reduced form equations or just any linear regression type of problem.

Simple Linear Relations

Let us consider a simple consumption function such as

$$C_t = \beta_0 + \beta_1 Z_t + u_t \qquad (18.8)$$

where C_t is quantity of consumption in t.

Z_t is amount of income in t.

u_t is a random disturbance

β_0 and β_1 are unknown parameters.

Suppose we have n observations on C and Z.

Assuming the economy conforms to the model and the true values of the parameters are β_0 and β_1, the probability of observing the particular sample $C_1 C_2 - - - C_n$ and $Z_1, Z_2, - - - Z_n$ is equal to the probability that the error term shall have assumed the particular values $u_1, u_2 - - - u_n$ multiplied by a factor $|J|$. $|J| = 1$ in all single equation cases regardless of the number of exogenous variables and we do not have to concern ourselves with it in single equation cases.

$|J|$ is a Jacobian which is the determinant of the matrix of partial derivatives in the multiple equation case and a 1 x 1 determinant in the single equation case. It arises when one is performing a transformation from one variable or set of variables to another variable or set of variables. In the transformation process, the differentials also have to be transformed or the appropriate differentials of the new variable or variables have to be determined. These give rise to the Jacobian.

The likelihood function corresponding to (18.8) is:

(18.9) $L = f(C_1 C_2 - - - C_n, Z_1 Z_2 - - - Z_n; \beta_0 \beta_1)$

which entails d C. But,

(18.10) $u_t = C_t - \beta_0 - \beta_1 Z_t$

and the likelihood function of interest is:

(18.11) $L = f(u_1, u_2, - - - u_n)$

which entails d_u. The Jacobian of interest is:

(18.12) $\frac{d u_t}{d C_t} = 1$ which is obtained directly from (18.10). It is clear that (18.12) would hold even if there was more than one Z variable.

In a system of equations where there is more than one endogenous variable, the Jacobian is the determinant of a matrix of first order partial derivatives and $|J| \neq 1$. In simultaneous systems the Jacobians are extremely important and are partly responsible for the complexity and

difficulty in computation of the maximum likelihood estimates of coefficients of simultaneous systems of linear equations.

Illustration of Jacobians

For purposes of illustrating a Jacobian that is not equal to one consider the following functions:

(18.13) $X = g(Z,W)$

(18.14) $Y = h(Z,W)$

The probability density is:

(18.15) $f(X,Y)\, dXdY$

and

(18.16) $f[g(Z,W), h(Z,W)] \quad |J| dZdW$

where $|J|$ is the Jacobian of transformation.

(18.17) $$|J| = \begin{vmatrix} \frac{\partial X}{\partial Z} & \frac{\partial X}{\partial W} \\ \frac{\partial Y}{\partial Z} & \frac{\partial Y}{\partial W} \end{vmatrix}$$

and

(18.18) $$|J| = \begin{vmatrix} \frac{\partial X}{\partial Z} & \frac{\partial X}{\partial W} \\ \frac{\partial Y}{\partial Z} & \frac{\partial Y}{\partial W} \end{vmatrix} = \begin{vmatrix} \frac{\partial X}{\partial Z} & \frac{\partial Y}{\partial Z} \\ \frac{\partial X}{\partial W} & \frac{\partial Y}{\partial W} \end{vmatrix} = |J'|$$

Single Equation Likelihood Function

Now returning to our linear consumption function, let us assume that the distribution of the disturbances, u_t's is normal with

(18.19) $E(u_t) = 0$

(18.20) $E(u_t u_t) = \sigma u_t u_t = \sigma^2$ for $t=1,2, - - - n$

and

(18.21) $E(u_t u_s) = \sigma u_t u_s = 0$ for $t \neq s$

The specific likelihood function is:

$$(18.22)\quad L(u_1 u_2 - - - u_n) = \frac{1}{\sqrt{2\P\sigma u_1 u_1}}\, e^{\frac{-[u_1-E(u_1)]^2}{2\sigma u_1 u_1}} \cdot \frac{1}{\sqrt{2\P\sigma u_2 u_2}}$$
$$e^{\frac{-[u_2-E(u_2)]^2}{2\sigma u_2 u_2}} - - - - - - \frac{1}{\sqrt{2\P\sigma u_n u_n}}\, e^{\frac{-[u_n-E(u_n)]^2}{2\sigma u_n u_n}}$$
$$= (2\P)^{\frac{-n}{2}} \left(\frac{1}{\sqrt{\sigma u_1 u_1}} \cdot \frac{1}{\sqrt{\sigma u_2 u_2}} - - - - - \cdot \frac{1}{\sqrt{\sigma u_n u_n}}\right) \cdot$$
$$\exp\left\{\frac{-1}{2}\left[\frac{[u_1-E(u_1)]^2}{\sigma u_1 u_1} + \frac{[u_2-E(u_2)]^2}{\sigma u_2 u_2} + - - - - - - +\right.\right.$$
$$\left.\left.\frac{[u_n-E(u_n)]^2}{\sigma u_n u_n}\right]\right\}$$

But considering (18.19) - (18.21) the likelihood function simplifies to:

$$(18.23)\quad L = (2\P)^{\frac{-n}{2}}\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left\{\frac{-1}{2}\left[\frac{u_1^2}{\sigma^2} + \frac{u_2^2}{\sigma^2} + - - - + \frac{u_n^2}{\sigma^2}\right]\right\}$$
$$= (2\P)^{\frac{-n}{2}}(\sigma^2)^{\frac{-n}{2}} \exp\left\{\frac{-1}{2\sigma^2}\sum_{t=1}^{n} u_t^2\right\}$$

Now (18.23) is the function we want to maximize. Let's substitute (18.10) into (18.23) and then use the logarithmic form of the likelihood function:

$$(18.24)\quad L = (2\P)^{\frac{-n}{2}}(\sigma^2)^{\frac{-n}{2}} \exp\left\{\frac{-1}{2\sigma^2}\sum_{t=1}^{n}(C_t - \beta_0 - \beta_1 Z_t)^2\right\}$$

and

$$(18.25)\quad \log L = \frac{-n}{2}\log(2\P) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(C_t - \beta_0 - \beta_1 Z_t)^2$$

Remember that we have a sample of n observations of C and Z and β_0 and β_1 are the unknowns. We are viewing β_0 and β_1 as variables. By taking first partial derivatives of log L with respect to each and setting these partial derivatives equal to zero we can solve for the values of β_0 and β_1 which maximize log L. These are the same values that

maximize L. Hence, for β_0

$$(18.26)\quad \frac{\partial \log L}{\partial \beta_0} = \frac{-1}{\partial \sigma^2}(2)(-1)\sum_{t=1}^{n}(C_t - \beta_0 - \beta_1 Z_t) = 0$$

and

$$(18.27)\quad \sum_{t=1}^{n} C_t - n\beta_0 - \beta_1 \sum_{t=1}^{n} Z_t = 0$$

$$n\beta_0 = \sum_{t=1}^{n} C_t - \beta_1 \sum_{t=1}^{n} Z_t =$$

thus,

$$(18.28)\quad \beta_0 = \bar{C} - \beta_1 \bar{Z}$$

With respect to β_1,

$$(18.29)\quad \frac{\partial \log L}{\partial \beta_1} = \frac{-1}{2\sigma^2}(2)(-Z_t)\sum_{t=1}^{n}(C_t - \beta_0 - \beta_1 Z_t) = 0$$

and

$$(18.30)\quad \sum_{t=1}^{n}(Z_t C_t - \beta_0 Z_t - \beta_1 Z_t^2) = 0$$

and

$$\sum_{t=1}^{n} Z_t C_t - \beta_0 \sum_{t=1}^{n} Z_t - \beta_1 \sum_{t=1}^{n} Z_t^2 = 0$$

now substituting the expression for β_0 in (18.28) into (18.30) we have:

$$(18.31)\quad \sum_{t=1}^{n} Z_t C_t - \left(\frac{\sum_{t=1}^{n} C_t - \beta_1 \sum_{t=1}^{n} Z_t}{n}\right)\sum_{t=1}^{n} Z_t - \beta_1 \sum_{t=1}^{n} Z_t^2 = 0$$

$$\sum_{t=1}^{n} Z_t C_t - \frac{\left(\sum_{t=1}^{n} C_t\right)\left(\sum_{t=1}^{n} Z_t\right)}{n} - \beta_1 \sum_{t=1}^{n} Z_t^2 + \beta_1 \frac{\left(\sum_{t=1}^{n} Z_t\right)^2}{n} = 0$$

Hence,

$$(18.32)\quad \beta_1\left[\sum_{t=1}^{n} Z_t^2 - \frac{\left(\sum_{t=1}^{n} Z_t\right)^2}{n}\right] = \sum_{t=1}^{n} Z_t C_t - \frac{\left(\sum_{t=1}^{n} C_t\right)\left(\sum_{t=1}^{n} Z_t\right)}{n}$$

and

$$(18.33)\quad \beta_1 = \frac{\sum_{t=1}^{n} Z_t C_t - \frac{\left(\sum_{t-1}^{n} C_t\right)\left(\sum_{t=1}^{n} Z_t\right)}{n}}{\sum_{t=1}^{n} Z_t^2 - \frac{\left(\sum_{t=1}^{n} Z_t\right)^2}{n}} = \frac{\sum_{t=1}^{n} z_t c_t}{\sum_{t=1}^{n} z_t^2} = \frac{m_{zc}}{m_{zz}}$$

where z_t and c_t are deviations about the means and m_{zc} and m_{zz} represent moments about the means.

The maximum likelihood estimators of the coefficients of the single linear consumption function given in (18.8) are:

$$(18.34)\quad \hat{\beta}_0 = \bar{C} - \hat{\beta}_1 \bar{Z}$$

and

$$(18.35)\quad \hat{\beta}_1 = \frac{\sum_{t=1}^{n} z_t c_t}{\sum_{t=1}^{n} z_t^2}$$

Note that $\hat{\beta}_0$ and $\hat{\beta}_1$ are the same as least squares estimators.

To obtain the maximum estimator of the variance of u we differentiate (18.25) with respect to σ^2 and solve for the value of σ^2 that will meximize log L.

Hence,

$$(18.36)\quad \frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2(\sigma^2)^2} - (-1)\frac{1}{2\sigma^2}\left[\sum_{t=1}^{n}(C_t - \beta_0 - \beta_1 Z_t)^2\right] = 0$$

and

$$(18.37)\quad \frac{n}{2\sigma^2} = \frac{1}{2(\sigma^2)^2}\left[\sum_{t=1}^{n}(C_t - \beta_0 - \beta_1 Z_t)^2\right]$$

and

$$(18.38)\quad \sigma^2 = \frac{\sum_{t=1}^{n}(C_t - \beta_0 - \beta_1 Z_t)^2}{n} = \frac{\sum_{t=1}^{n} u_t^2}{n}$$

The maximum likelihood estimators of the variance of u is:

$$(18.39)\quad \tilde{\sigma}^2 = \frac{\sum_{t=1}^{n}(C_t - \tilde{\beta}_0 - \tilde{\beta}_1 Z_t)^2}{n} = \frac{\sum_{t=1}^{n} \tilde{u}_t^2}{n}$$

Notice that in $\tilde{\sigma}^2$ the loss of degrees of freedom are not taken into consideration and the estimator is biased from the viewpoint of a small sample but asymptotically unbiased.

An unbiased estimator of the variance of u may be readily obtained by utilizing information about the distribution of the ratio of a sum of

a sum of squares over a true variance. Such a ratio is by definition a χ^2 (Chi-square). $\frac{n\tilde{\sigma}^2}{\sigma^2}$ distributes itself as a χ^2 with (n-2) degrees of freedom in this case and the distribution is independent of $\tilde{\beta}_0$ and $\tilde{\beta}_1$. Thus,

$$(18.40) \quad \chi^2 = \frac{(n-2)\hat{\sigma}^2}{\sigma^2}$$

where $\hat{\sigma}^2$ is an unbiased estimator of the variance of u and

$$(18.41) \quad \frac{n\tilde{\sigma}^2}{\sigma^2} = \frac{(n-2)\hat{\sigma}^2}{\sigma^2}$$

which yields

$$(18.42) \quad \tilde{\sigma}^2 = \frac{(n-2)\hat{\sigma}^2}{n}$$

now substituting (18.42) into (18.39) we have

$$(18.43) \quad \frac{(n-2)\hat{\sigma}^2}{n} = \frac{\sum_{t=1}^{n}(C_t - \tilde{\beta}_0 - \tilde{\beta}_1 Z_t)^2}{n}$$

and

$$(18.44) \quad \hat{\sigma}^2 = \frac{\sum_{t=1}^{n}(C_t - \tilde{\beta}_0 - \tilde{\beta}_1 Z_t)^2}{n-2}$$

where $\hat{\sigma}^2$ is an unbiased estimator of the variance of u and is the same as the least squares estimator of the residual variance.

In summarizing the use of maximum likelihood estimation for coefficients of a simple linear relationship the estimators of the coefficients are unbiased but the maximum likelihood estimator of the variance of u is biased. An unbiased estimator of the variance of u may be readily obtained by utilizing information that the ratio of a sum of squares over a true variance distributes itself as a χ^2 with degrees of freedom which take into consideration the loss of degrees of freedom due to imposed restrictions.

Multiple Variable Linear Relations

Let us briefly take a look at maximum likelihood estimation for a multiple variable linear relationship. Consider the following relationship:

(18.45) $Y_t = \beta_0 + \beta_1 Z_{1t} + \beta_2 Z_{2t} + - - - - - + \beta_k Z_{kt} + u_t$

Suppose we have a sample of n observations for Y_t and each of the Z's.

Equation (18.45) may be expressed in matrix notation as follows:

(18.46) $Y = Z\beta + u$

where Y is n x 1

Z is n x (k+1)

β is (k+1) x 1

u is n x 1

and

(18.47) $u = Y - Z\beta$

Disregarding a lot of "in between details" we can formulate the following likelihood function:

(18.48) $L = (2\pi\sigma^2)^{\frac{-n}{2}} \exp\left\{\frac{-u'u}{2\sigma^2}\right\}$

assuming as before that:

(18.49) $E(u_t) = 0$

(18.50) $E(u'u) = \sigma u_t u_t = \sigma^2$ for t=1, 2, - - - n

and

(18.51) $E(u_t u_s) = \sigma_{u_t u_s} = 0$ for $t \neq s$

The logarithmic form of the likelihood function is:

(18.52) $\log L = \frac{-n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 \frac{-u'u}{2\sigma^2}$

Substituting (18.47) into (18.52) yields:

(18.53) $\log L = \frac{-n}{2} \log (2\pi) \frac{-n}{2} \log \sigma^2 \frac{-(Y - Z\beta)'(Y - Z\beta)}{2\sigma^2}$

Now by partially differentiating (18.53) with respect to the β vector, setting these partial derivatives equal to zero we can solve for the values of β that will maximize log L and thus L. Hence,

(18.54) $\frac{\partial \log L}{\partial \beta} = \frac{-1}{2\sigma^2} [-2Z'(Y-Z\beta)] = 0$

and denoting the parameters to be estimated with a curl:

(18.55) $Z'Y - (Z'Z)\hat{\beta} = 0$

$$(Z'Z)\hat{\beta} = Z'Y$$

Thus,

(18.56) $\hat{\beta} = (Z'Z)^{-1}Z'Y$

which is the same as the least squares estimator of β.

In a similar way we can solve for σ^2.

(18.57) $$\frac{\partial \log L}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{(Y-Z\beta)'(Y-Z\beta)}{2(\sigma^2)^2}$$

Denoting the maximum estimator with a curl and setting (18.57) equal to zero we have:

(18.58) $$\frac{-n}{2\tilde{\sigma}^2} + \frac{(Y-Z\hat{\beta})'(Y-Z\hat{\beta})}{2(\tilde{\sigma}^2)^2} = 0$$

Solving for the variance estimator we have

(18.59) $$\tilde{\sigma}^2 = \frac{(Y-Z\hat{\beta})'(Y-Z\hat{\beta})}{n}$$

which is a biased estimator of the residual variance. Utilizing the information that

(18.60) $\frac{n\tilde{\sigma}^2}{\sigma^2} = \chi^2$ with (n-k-1) degrees of freedom and going through the same steps as indicated in (18.41) - (18.44) we can obtain the unbiased estimator:

(18.61) $$\hat{\sigma}^2 = \frac{(Y-Z\hat{\beta})'(Y-Z\hat{\beta})}{n-k-1}$$

In developing the estimators of the multiple variable case, we made the same assumptions as we did for the simple case plus that (Z'Z) is of full rank and that rank is less than or equal to (k+1). These are the same assumptions made in obtaining least squares estimators plus we assume the disturbances to be normally distributed in order to obtain maximum likelihood estimators.

Simultaneous Linear Equations

In this section we will address ourselves to applying maximum likelihood estimation to a simultaneous system of linear equations.

Attention will be devoted to formulating the appropriate liklihood function and specifically setting forth the appropriate Jacobian. Emphasis will be on showing how one goes about getting full information maximum likelihood estimators of coefficients of a simultaneous system of linear equations, but we will not carry through to actually getting the estimator. Why we do not carry this procedure through to getting the estimators will become apparent as difficulties of nonlinear equations in the unknown parameters are pointed out. These nonlinearities arise because of the Jacobian with which we are confronted in simultaneous systems of linear equations arising from the nonlinear nature of the partial derivatives of the quadratic forms of disturbances with which we have to deal.

Let us consider a four equation model where the four equations may be viewed as supply and demand equations of two commodities. β denotes parameters of endogenous variables and γ's denote parameters of predetermined variables, exogenous and lagged endogenous variables. The model is:

$$(18.62) \quad Y_{1t} = \beta_{12}Y_{2t} + \beta_{14}Y_{4t} + \gamma_{11}Z_{1t} + u_{1t}$$

$$(18.63) \quad Y_{2t} = \beta_{21}Y_{1t} + \gamma_{22}Z_{2t} + \gamma_{23}Z_{3t} + u_{2t}$$

$$(18.64) \quad Y_{3t} = \beta_{32}Y_{2t} + \beta_{34}Y_{4t} + \gamma_{31}Z_{1t} + u_{3t}$$

$$(18.65) \quad Y_{4t} = \beta_{43}Y_{3t} + \gamma_{42}Z_{2t} + \gamma_{44}Z_{4t} + u_{4t}$$

where: Y_1 is price of commodity 1.

Y_2 is quantity of commodity 1

Y_3 is price of commodity 2

Y_4 is quantity of commodity 2

Z_1 is income

Z_2 is price of factor 1 (labor)

Z_3 is price of factor 2 (land)

Z_4 is price of factor 3 (capital)

(18.62) and (18.64) may be viewed as demand relations while (18.63) and (18.65) may be viewed as representing supply relations. In each equation the endogenous variable upon which the equation has been normalized has a coefficient of one. This particular formulation is given mainly to more clearly illustrate how the Jacobian comes into being. Each of the equations (18.62)-(18.65) is overidentified according to the counting rule and a scrutiny prior to estimation of the necessary and sufficient condition indicates that for each equation there is at least one way the condition could be fulfilled if certain coefficients are not zero. The model is considered to be over-identified.

Now considering a sample of n observations for each of the Y's and Z's i.e., t=1, 2, - - -, n we are interested in determining the most likely values of β's and γ's to have generated the particular sample of observations.

The n values of u_1, u_2, u_3, and u_4 are jointly distributed as:

(18.66) $$\prod_{t=1}^{n} p(u_{1t}, u_{2t}, u_{3t}, u_{4t})\, du_{1t} du_{2t} du_{3t} du_{4t}$$

and the Y's follow the same distribution.

(18.67) $$\prod_{t=1}^{n} p(u_{1t}, u_{2t}, u_{3t}, u_{4t}) \left| \frac{\partial(u_{1t}, u_{2t}, u_{3t}, u_{4t})}{\partial(Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t})} \right| dY_{1t} dY_{2t} dY_{3t} dY_{4t}$$

$\left|\frac{\partial(u_{1t}, u_{2t}, u_{3t}, u_{4t})}{\partial(Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t})}\right|$ is the Jacobian which we will explore in more detail later.

We will assume that the u's are <u>normally distributed with zero means and constant variance</u>. The variance-covariance matrix of the disturbances is:

$$(18.68)\quad \Phi = \begin{bmatrix} E(u_1^2) & E(u_1u_2) & E(u_1u_3) & E(u_1u_4) \\ E(u_2u_1) & E(u_2^2) & E(u_2u_3) & E(u_2u_4) \\ E(u_3u_1) & E(u_3u_2) & E(u_3^2) & E(u_3u_4) \\ E(u_4u_1) & E(u_4u_2) & E(u_4u_3) & E(u_4^2) \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix}$$

where σ_{ij} is a variance when i=j and the covariance between the disturbance from equations i and j when $i \neq j$; i, j=1, 2, 3, 4.

One possible case which we can explore is when the variance-covariance matrix ϕ is a diagonal matrix.

$$(18.69)\quad \phi = \begin{bmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{bmatrix}$$

This occurs when $E(u_iu_j) = 0$ if $i \neq j$; i,j = 1, 2, 3, 4. Whenever the variance-covariance matrix of disturbances among equations is a diagonal matrix, the disturbances are not jointly distributed and hence, the endogenous variables are not jointly distributed and thus not jointly determined. This means the system of equations under consideration is really a system of equations which can be handled one at a time separately and independently of each other. Cranking through the maximum likelihood estimators of each of the coefficients of the system would have resulted in least squares estimators for each. Recursive models and sets of independent regression equations exhibit a diagonal variance-covariance matrix. Simultaneity exists only if some of the covariances are not zero.

Considering the simultaneous equation case where u_1, u_2, u_3, and u_4 are jointly distributed the likelihood function is based on a

multivariate normal distribution. In order to simplify the expression of the likelihood function the disturbance need to be expressed in matrix form

(18.70) $u = BY + \Gamma Z$

where

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}$$

The likelihood function is:

$$(18.71)\quad L = |J|^n \left[\frac{1}{(2\pi)^{\frac{1}{2}}}\right]^{4n} \frac{1}{|\phi|^{\frac{n}{2}}} e^{-\frac{1}{2}u'\phi^{-1}u}$$

$$= |J|^n (2\pi)^{\frac{-4n}{2}} |\phi|^{\frac{-n}{2}} e^{-\frac{1}{2}(BY + \Gamma Z)'\phi^{-1}(BY + \Gamma Z)}$$

and the logarithm of the likelihood function is:

$$(18.72)\quad \log L = n \log|J| - \frac{4n}{2}\log(2\pi) - \frac{n}{2}\log|\phi| - \frac{1}{2}(BY + \Gamma Z)'\phi^{-1}(BY + \Gamma Z)$$

where

$$(18.73)\quad \phi^{-1} = \begin{bmatrix} \sigma^{11} & \sigma^{12} & \sigma^{13} & \sigma^{14} \\ \sigma^{21} & \sigma^{22} & \sigma^{23} & \sigma^{24} \\ \sigma^{31} & \sigma^{32} & \sigma^{33} & \sigma^{34} \\ \sigma^{41} & \sigma^{42} & \sigma^{43} & \sigma^{44} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix}^{-1}$$

and σ^{ij} is the ijth element of the inverse of the variance-covariance matrix of disturbances.

At this point it may appear that the maximization of (18.72) is quite straight forward and it is, but the work involved is quite burdensome. First, maximization is with respect to all the coefficients, variances, and covariances. Some complex and difficult problems are disguised in (18.72). These will become apparent as we specify the specific elements of $|J|$ and attempt to get estimated values of the elements of ϕ. Lets start with the Jacobian.

$$(18.74)\qquad |J| = \left|\frac{\partial(u_{1t}, u_{2t}, u_{3t}, u_{4t})}{\partial(Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t})}\right| = \begin{vmatrix} \frac{\partial u_1}{\partial Y_1} & \frac{\partial u_1}{\partial Y_2} & \frac{\partial u_1}{\partial Y_3} & \frac{\partial u_1}{\partial Y_4} \\ \frac{\partial u_2}{\partial Y_1} & \frac{\partial u_2}{\partial Y_2} & \frac{\partial u_2}{\partial Y_3} & \frac{\partial u_2}{\partial Y_4} \\ \frac{\partial u_3}{\partial Y_1} & \frac{\partial u_3}{\partial Y_2} & \frac{\partial u_3}{\partial Y_3} & \frac{\partial u_3}{\partial Y_4} \\ \frac{\partial u_4}{\partial Y_1} & \frac{\partial u_4}{\partial Y_2} & \frac{\partial u_4}{\partial Y_3} & \frac{\partial u_4}{\partial Y_4} \end{vmatrix}$$

In order to obtain the Jacobian we need to rewrite equations (18.62)-(18.65) in the following form:

$$(18.75)\qquad u_{1t} = Y_{1t} - \beta_{12}Y_{2t} - \beta_{14}Y_{4t} - \gamma_{11}Z_{1t}$$

$$(18.76)\qquad u_{2t} = Y_{2t} - \beta_{21}Y_{1t} - \gamma_{22}Z_{2t} - \gamma_{23}Z_{3t}$$

$$(18.77)\qquad u_{3t} = Y_{3t} - \beta_{32}Y_{2t} - \beta_{34}Y_{4t} - \gamma_{31}Z_{1t}$$

$$(18.78)\qquad u_{4t} = Y_{4t} - \beta_{43}Y_{3t} - \gamma_{42}Y_{3t} - \gamma_{44}Z_{4t}$$

Now,

$$(18.79)\qquad |J| = \begin{vmatrix} 1 & -\beta_{12} & 0 & -\beta_{14} \\ -\beta_{21} & 1 & 0 & 0 \\ 0 & -\beta_{32} & 1 & -\beta_{34} \\ 0 & 0 & -\beta_{43} & 1 \end{vmatrix}$$

$$= 1\begin{vmatrix} 1 & 0 & 0 \\ -\beta_{32} & 1 & -\beta_{31} \\ 0 & \beta & 1 \end{vmatrix} - \beta_{21}(-1)\begin{vmatrix} -\beta_{12} & 0 & -\beta_{14} \\ -\beta_{32} & 1 & -\beta_{34} \\ 0 & -\beta_{43} & 1 \end{vmatrix}$$

$$= 1 - \beta_{34}\beta_{43} + \beta_{21}(-\beta_{12})\begin{vmatrix} 1 & -\beta_{34} \\ -\beta_{43} & 1 \end{vmatrix} + \beta_{21}\beta_{32}\begin{vmatrix} 0 & -\beta_{14} \\ -\beta_{43} & 1 \end{vmatrix}$$

$$= 1 - \beta_{34}\beta_{43} - \beta_{21}\beta_{12}(1-\beta_{34}\beta_{43}) + \beta_{21}\beta_{32}(-\beta_{14}\beta_{43})$$

$$|J| = 1 - \beta_{34}\beta_{43} - \beta_{21}\beta_{12} + \beta_{21}\beta_{12}\beta_{34}\beta_{43} - \beta_{21}\beta_{32}\beta_{14}\beta_{43}$$

Note that $|J|$ is really $|\beta|$.

But these are only part of the difficulties. The $u'u$ terms in (18.71) will bring in products of parameters in different equations where they did not exist before. This is clear from (18.75)-(18.78). Furthermore, we need to obtain estimates of the elements of ϕ in order to obtain $|\phi|$ and ϕ^{-1}. One thing which can be done is to replace the expectations in (18.69) with means over the sample observations of $\tilde{u}_i\tilde{u}_j$. This yields the following estimated variance-covariance matrix:

$$(18.80)\quad \tilde{\phi} = \frac{1}{n}\begin{bmatrix} \Sigma\tilde{u}_{1t}^2 & \Sigma\tilde{u}_{1t}\tilde{u}_{2t} & \Sigma\tilde{u}_{1t}\tilde{u}_{3t} & \Sigma\tilde{u}_{1t}\tilde{u}_{4t} \\ \Sigma\tilde{u}_{2t}\tilde{u}_{1t} & \Sigma\tilde{u}_{2t}^2 & \Sigma\tilde{u}_{2t}\tilde{u}_{3t} & \Sigma\tilde{u}_{2t}\tilde{u}_{4t} \\ \Sigma\tilde{u}_{3t}\tilde{u}_{1t} & \Sigma\tilde{u}_{3t}\tilde{u}_{2t} & \Sigma\tilde{u}_{3t}^2 & \Sigma\tilde{u}_{3t}\tilde{u}_{4t} \\ \Sigma\tilde{u}_{4t}\tilde{u}_{1t} & \Sigma\tilde{u}_{4t}\tilde{u}_{2t} & \Sigma\tilde{u}_{4t}\tilde{u}_{3t} & \Sigma\tilde{u}_{4t}^2 \end{bmatrix} \quad 2/$$

where the estimated values of the u_t's come from:

$$(18.81)\quad \tilde{u}_{1t} = Y_{1t} - \tilde{\beta}_{12}Y_{2t} - \tilde{\beta}_{14}Y_{4t} - \tilde{\gamma}_{11}Z_{1t}$$

$$(18.82)\quad \tilde{u}_{2t} = Y_{2t} - \tilde{\beta}_{21}Y_{1t} - \tilde{\gamma}_{22}Z_{2t} - \tilde{\gamma}_{23}Z_{3t}$$

$$(18.83)\quad \tilde{u}_{3t} = Y_{3t} - \tilde{\beta}_{32}Y_{2t} - \tilde{\beta}_{34}Y_{4t} - \tilde{\gamma}_{31}Z_{1t}$$

$$(18.84)\quad \tilde{u}_{4t} = Y_{4t} - \tilde{\beta}_{43}Y_{3t} - \tilde{\gamma}_{42}Z_{2t} - \tilde{\gamma}_{44}Z_{4t}$$

In the logarithmic form of the likelihood function, (18.72) we substitute $\tilde{\phi}$ given in (18.80) for ϕ and then obtain $\tilde{\phi}^{-1}$. $\tilde{\phi}$ is expressed in terms of Y's, Z's, $\tilde{\beta}$'s, and $\tilde{\gamma}$'s by expressing the $\tilde{u}$'s in terms of (18.81)-(18.84). At the same time replacing u_t' and u_t by $\tilde{u}_t'$ and $\tilde{u}_t$, log L may be written:

2/ All summations are from t = 1 to t = n.

$$(18.85)\quad \log L = n\log|J| - \frac{4n}{2}\log(2\pi) - \frac{n}{2}\log|\tilde{\phi}| - \frac{1}{2}(\tilde{\beta}Y + \tilde{\Gamma}Z)'\tilde{\phi}^{-1}(\tilde{\beta}Y + \tilde{\Gamma}Z)$$

The maximum likelihood estimators of the coefficients are obtained by differentiating (18.85) with respect to each of the coefficients, setting each of the resulting equations equal to zero and solving for the estimators. It should be apparent that this is not an easy task even for the simple four equation system considered here. There will be no attempt made to present these equations nor to show the expressions for estimators of specific β's and γ's. But focusing just on $|\tilde{\phi}|$ gives some feel of the complexity of the equations arising from the partial differentiation. Note if (18.81)-(18.84) are substituted into (18.80) that each of the elements of $\tilde{\phi}$ consists of an expression that is of second order with respect to the parameters. On the diagonal elements we know this will involve squared terms of individual parameters while off-diagonal elements will entail second order cross-products of parameters. Thus, in solving for the determinant of $\tilde{\phi}$ the determinental equation will contain terms that are of eight order in terms of the parameters. This doesn't mean that the expressions contain individual parameters to the eight power, but it does mean that the equations resulting from the partial differentiation that have to be solved involve expressions that are of eight order with respect to the particular combination of β's and γ's contained in the individual expressions. In addition the partial derivatives of the exponent lead to comparably difficult expressions. The value of $|J|$ depends directly on the parameters of the endogenous variables while the value of $|\phi|$ depends indirectly on all the parameters from equations (18.68)-(18.71).

Even though maximum likelihood estimators possess desirable statistical properties, the computational problems arising from the nonlinear equations which have to be solved are massive and expensive even with today's computer facilities. The problem is more of "software" than "hardware". For these reasons full information maximum likelihood is not very frequently used in estimating the coefficients of an over-identified simultaneous system of linear equations. Although some efficiency may be sacrificed, ease of computation makes other estimation procedures attractive.

CHAPTER 19

Derived Reduced Forms and Final Forms of Equations

Introduction

Multiple endogenous variables in the individual structural equations of simultaneous equation systems lead to some difficulties of interpretation of structural coefficients. When interpreting an individual coefficient it is untenable for all other variables in the equation to be non-variant. Values of other endogenous variables will vary simultaneously. Hence, there is a problem of trying to aasess the impacts of changes in variables using the estimated structural coefficients. The question which arises is are the estimated structural coefficients estimates of net or gross effects of changes in their respective variables and which type of information is needed to answer the problem or problems being analyzed.

In this chapter attention will be first devoted to developing in more detail the difficulties with interpreting estimated structural coefficients and the relation of these estimated coefficients to economic parameters such as price elasticities, cross-elasticities, income elasticities, etc. Then analytically drived reduced forms will be developed and the types of information provided by the derived coefficients of these forms will be discussed. Finally, considerable attention will be aimed at the notion of final forms of equations and at the various aspects of estimating and interpreting impact multipliers, interim multipliers, and total multipliers.

Difficulties Encountered in Interpreting Estimated Structural Coefficients 1/

To illustrate some of the difficulties of interpreting estimated structural coefficients from simultaneous equation systems, the concepts of

---

1/ This section draws heavily on Haidacher, R. C., and Penn, J. B. "Some Consideration in Interpreting Elasticities and Flexibilities From Si-Multaneous Systems" unpublished paper Economic Research Service, U. S. Department of Agriculture, February 1974.

elasticities and flexibilities will be drawn upon. Elasticities and flexibilities are estimated or computed in many econometric analyses and are frequently used to assess the implications of such analyses or studies. Empirical estimates of elasticities or flexibilities are often major inputs into analyses of proposed policy alternatives where price and/or income responsiveness are important. This would include most supply, demand, and price analysis studies and many studies concerned with capital, credit, and finance.

One of the major sources of confusion concerning elasticities and flexibilities has to do with their meaning and measurement. Using the concepts when certain crucial conditions can't be fulfilled has been partly responsible for some of the difficulty. For example using the concept of elasticity, which is applicable in a simple Marshallion context, in a more complex general framework such as the Hicksion framework of individual consumer behavior. Furthermore,to make such models more compatible with real world behavior then simultaneity is explicitly recognized and multiple jointly determined variables are introduced into each equation of the framework. The consequence is that the ceteris paribus conditions needed to interpret individual coefficients and elasticities in the Marshallion context are untenable in the more complex frameworks.

To illustrate the difficulty of interpretation of the estimated structural coefficients let us use elasticities as the economic parameters of interest. The concept of elasticity in economics is attributable to Alfred Marshall. Marshall defined elasticity of demand as the percentage change in quantity divided by the percentage change in price, when the percentage change in price is small. This definition assumes that some types of market demand relation exists in which quantity (Q) is some function of price (P), other things non-variant. Consider the following demand relation:

(19.1) $Q = f(P)$

In terms of partial derivatives the price elasticity of demand defined by Marshall is:

$$(19.2) \quad E = \frac{\partial Q}{\partial P} \cdot \frac{P}{Q}$$

Writing the inverse function of (19.1) as:

$$(19.3) \quad P = f^{-1}(Q)$$

the price flexibility is defined as:

$$(19.4) \quad F = \frac{\partial P}{\partial Q} \cdot \frac{Q}{P} = \frac{1}{E}$$

The price flexibility as expressed in (19.4) is the percentage change in price divided by the percentage chnage in quantity when the percentage change in quantity is small ceteris paribus.

Single Equation Models

Single equation models where a single endogenous variable is some function of a set of exogenous variables cause no problem with regard to interpretation of coefficients or with regard to elasticities or flexibilities derived from such estimed coefficients. For example, consider the following relation:

$$(19.5) \quad Q = f(Z_1, Z_2, Z_3, - - - Z_n, \varepsilon)$$

where the Q is the quantity demanded, the Z's are exogenous vatiables, and the ε is a random distrubance term. An elasticity with respect to any of the exogenous variables can be obtained as:

$$(19.6) \quad E_i = \frac{\partial Q}{\partial Z_i} \cdot \frac{Z_i}{Q}$$

The basic assumptions which allow (19.6) to be interpreted in a straight forward way as an elasticity are that the quantity is endogenous and functionally related to the Z variables and that the responsiveness of quantity to a change in any $i^{\underline{th}}$ exogenous variable may be interpreted individually with all other exogenous variables non-variant. Furthermore, it is being assumed that the supply relation and other relations which might make up the model are behaving

in such a manner that prices are predetermined.

Simultaneous Equations Model

Now let us consider a simultaneous system of linear equations consisting of two endogenous variables and two equations. Consider the first equation in the following model a demand relation and the second a supply relation:

$$\beta_{11}Y_1 + \beta_{12}Y_2 + \gamma_{11}Z_1 + \gamma_{12}Z_2 + \gamma_{10}Z_0 = u_1 \tag{19.7}$$

$$\beta_{21}Y_1 + \beta_{22}Y_2 + \gamma_{23}Z_3 + \gamma_{24}Z_4 + \gamma_{20}Z_0 = u_2$$

The Y's are endogenous variables, the Z's are exogenous and the u's are random disturbances. The variables are:

Y_1 = quantity demanded = quantity supplied.

Y_2 = own price

Z_1 = price of a substitute

Z_2 = income

Z_3 = price of an input

Z_4 = a measure of weather

Z_0 = a vector of ones.

β's are parameters of the endogenous variables.

γ's are parameters of the exogenous variables.

u's are random jointly and normally distributed disturbances with means

$u_1 = u_2 = 0$.

In matrix for (19.7) may be written as:

$$\begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 & 0 & \gamma_{10} \\ 0 & 0 & \gamma_{23} & \gamma_{24} & \gamma_{20} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_0 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{19.8}$$

Great care needs to be exercised in interpreting the estimated coefficients of (19.7) or (19.8) and elasticities or flexibilities that might be derived from such estimated coefficients. The presence of more than one endogenous variable in each euqation raises serious questions about the validity of partial differentiation since the ceteris paribus assumption is violated and hence the traditional definition of elasticity or flexibility is violated. If one of the exogenous variables changes, than all jointly determined variables will be affected which means that not only the endogenous variable under consideration will be affected but other endogenous variables in the system will be affected and changes in these other endogenous variables will in turn affect the endogenous variable under consideration. Furthermore, if one equation of a simultaneous system can legitimately be considered in isolation then specifying the system as a simultaneous equation system is inappropriate and questionable. A simultaneous equation specification is not compatible with analyzing individual equations separately and independently of other equations in the system particularly if the focus of analyses is on assessing the net inputs of changes in exogenous variables.

To further illustrate the difficulties of interpretation being emphasized, let us normalize the first equation if (19.7) with respect to Y_1 and the second with respect to Y_2. Let normalization be $\beta_{11} = \beta_{22} = -1$. The resulting normalized set of equations is:

$$
\begin{aligned}
Y_1 &= \beta_{12}Y_2 + \gamma_{11}Z_1 + \gamma_{12}Z_2 + \gamma_{10}Z - u_1 \\
Y_2 &= \beta_{21}Y_1 + \gamma_{23}Z_3 + \gamma_{24}Z_4 + \gamma_{20}Z_0 - u_2
\end{aligned}
\tag{19.9}
$$

Assume that the parameters have been estimated by a suitable estimation technique such as two-stage least squares and the estimated relationships are

$$
\begin{aligned}
Y_1 &= \hat{\beta}_{12}Y_2 + \hat{\gamma}_{11}Z_1 + \hat{\gamma}_{12}Z_2 + \hat{\gamma}_{10} - \hat{u}_1 \\
Y_2 &= \hat{\beta}_{21}Y_1 + \hat{\gamma}_{23}Z_3 + \hat{\gamma}_{24}Z_4 + \hat{\gamma}_{20} - \hat{u}_2
\end{aligned}
\tag{19.10}
$$

Now we can clearly see the possible difficulty in interpreting the impact of a change in one of the exogenous Z's on one of the endogenous Y's. For illustrative purposes suppose Z_1 changes by some incremental amount dZ_1. By how much does Y_1 change? An important issue is are we interested in the net amount or the gross amount by which Y_1 will change. From (19.10) we see that for every unit change in Z_1 the corresponding change in Y_1 is $\hat{\gamma}_{11}$. But this is the gross change in Y_1 associated with a unit change in Z_1 all other exogenous variables non-variant. From the second equation in (19.10) it is obvious that a change in Y_1 simultaneously leads to a change in Y_2. Every unit change in Y_1 leads to a $\hat{\beta}_{21}$ simultaneous change in Y_2 and changes in Y_2 lead to simultaneous changes in Y_1. From the first equation in (19.10) the simultaneous change in Y_1 associated with a unit change in Y_2 is $\hat{\beta}_{12}$. Clearly to get at the net effect of a change in Z, the simultaneous changes among the endogenous variables have to be taken into consideration. Now the net affect of a unit change in Z_1 will be something greater than or less than $\hat{\gamma}_{11}$ depending on the signs and magnitudes of $\hat{\beta}_{21}$ and $\hat{\beta}_{12}$. Also, if interest is in the net effects of changes in exogenous variables then the form of equations as presented in (19.10) are not the most amenable to analyzing such net effects. Furthermore, elasticities derived from the estimated coefficients of (19.10) need to be interpreted with great caution because the ceterus paribus assumptions can not be met. At best they should be interpreted in terms of relative gross changes. The traditional definition of elasticity refers to relative net changes but it is not possible to obtain these from the structural equations of the form (19.10)

Analytically Derived Reduced Forms

The net impacts of changes in exogenous variables on endogenous variables may be conveniently assessed from the analytically derived reduced forms. Also, these forms may be used for predicting the values of endogenous variables since each endogenous variable is expressed as a function of exogenous variables. The analytically derived reduced forms are obtained by solving (19.8) for the Y's. The analytically derived reduced forms for the model being considered are:

$$(19.11)\quad \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = -\begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}^{-1} \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 & 0 & 0 \\ 0 & 0 & \gamma_{23} & \gamma_{24} & \gamma_{20} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_0 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}^{-1} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

where:

$$(19.12)\quad \beta^{-1} = \frac{1}{\beta_{11}\beta_{22} - \beta_{12}\beta_{21}} \begin{bmatrix} \beta_{22} & -\beta_{12} \\ -\beta_{21} & \beta_{11} \end{bmatrix}$$

and hence,

$$(19.13)\quad \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = -\frac{1}{\beta_{11}\beta_{22}-\beta_{12}\beta_{21}} \begin{bmatrix} \beta_{22}\gamma_{11} & \beta_{22}\gamma_{12} & -\beta_{12}\gamma_{23} & -\beta_{12}\gamma_{24} \\ -\beta_{21}\gamma_{11} & -\beta_{21}\gamma_{12} & \beta_{11}\gamma_{23} & \beta_{11}\gamma_{24} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix}$$

$$\begin{bmatrix} (\beta_{22}\gamma_{10} & -\beta_{12}\gamma_{20}) \\ (-\beta_{21}\gamma_{10} & +\beta_{11}\gamma_{20}) \end{bmatrix} \begin{bmatrix} Z_0 \end{bmatrix}$$

$$+ \frac{1}{\beta_{11}\beta_{22}-\beta_{12}\beta_{21}} \begin{bmatrix} \beta_{22} & -\beta_{12} \\ -\beta_{21} & \beta_{11} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Thus,

$$(19.14)\quad Y_1 = -\frac{1}{\beta_{11}\beta_{22} - \beta_{12}\beta_{21}} [\beta_{22}\gamma_{11}Z_1 + \beta_{22}\gamma_{12}Z_2 - \beta_{12}\gamma_{23}Z_3 - \beta_{12}\gamma_{24}Z_4 + (\beta_{22}\gamma_{10} - \beta_{12}\gamma_{20})Z_0] + \frac{\beta_{22}u_1 - \beta_{12}u_2}{\beta_{11}\beta_{22} - \beta_{12}\beta_{21}}$$

and

$$(19.15)\quad Y_2 = -\frac{1}{\beta_{11}\beta_{22} - \beta_{12}\beta_{21}} [-\beta_{21}\gamma_{11}Z_1 - \beta_{21}\gamma_{12}Z_2 + \beta_{11}\gamma_{23}Z_3 + \beta_{11}\gamma_{24}Z_4 + (\beta_{11}\gamma_{20} - \beta_{21}\gamma_{10})Z_0] + \frac{\beta_{11}u_2 - \beta_{21}u_1}{\beta_{11}\beta_{22} - \beta_{12}\beta_{21}}$$

Depending on normalization of the equations some of the β's would be 1 or -1. For example, if the model was normalized as in (19.9), $\beta_{11} = \beta_{22} = 1$. Now from (19.14) we can see that the net impact on Y_1 of a unit change in Z_1, is:

$$(19.16)\quad \frac{-\beta_{22}\gamma_{11}}{\beta_{11}\beta_{22} - \beta_{12}\beta_{21}}$$

In a similar manner the net impacts of changes in any of the exogenous variables, individually or in combination, may be evaluated. These net impacts are frequently the ones sought by policy makers. Also (19.14) and (19.15) may be used to predict Y_1 and Y_2 respectively as long as there is some way available of obtaining the values of the exogenous variables.

Furthermore, now if we wish to derive certain elasticity concerned with net impacts, we are in a position to do so. For example, if we were interested in the elasticity of Y_1 with respect to Z_1 we can utilize (19.16) since:

$$(19.17)\quad E_{Y_1,Z_1} = \frac{\partial Y_1}{\partial Z_1} \cdot \frac{Z_1}{Y_1} = \frac{-\beta_{22}\gamma_{11}}{\beta_{11}\beta_{22} - \beta_{12}\beta_{21}} \left(\frac{Z_1}{Y_1}\right)$$

Also, if we were interested in the elasticity of one endogenous variable with respect to another, in this case say Y_1 with respect to Y_2, we could utilize the ratio of the respective differentials. In the particular case of the elasticity of Y_1 with respect to Y_2 we have:

$$(19.18)\quad E_{Y_1,Y_2} = \frac{dY_1}{dY_2} \cdot \frac{Y_2}{Y_1}$$

The differentials of the endogenous variables could arise because of changes in one or more exogenous variables and the particular exogenous variables that change to yield dY_1 are the same ones that change to yield dY_2. Given a simultaneous equation system, changes in endogenous variables can occur only if one or more exogenous variables change.

Final Form of Econometric Equation Systems

The model specified in (19.7) with which we have been dealing is a simple one and all variables pertain to the time period t. In much policy related research interest is in assessing the net impacts of changes in exogenous variables on selected endogenous variables. Furthermore, the interest frequently is in assessing the impacts not only in a single time period but in evaluating the accumulated impacts for a specified amount of elapsed time and sometimes the total accumulated impacts may be of interest. Also, for the purpose of evaluating policies an analysis of the time path of the response of an endogenous variable to changes in one or more exogenous variables during a single time period may be of interest. The need for this type of information has to be taken into consideration in the initial phases of modelling so that the appropriate lagged endogenous and exogenous variables are incorporated into the model. Information about the various types of impact multipliers may be obtained from the analytically derived reduced forms if the structural model was formulated with appropriate lagged endogenous and exogenous variables to permit the impacts of changes in exogenous variables to be assessed for any desired time period. One of the extremely fine discussions of the modelling, estimation, and interpretation of final equation forms is given by

Theil and Boot.[2/]

A point of departure for discussing final equation forms and impact multipliers is a system of analytically derived reduced form equations:

$$(19.19)\quad Y_t = AY_{t-1} + BX_t + CX_{t-1} + u_t^*$$

where

Y_t is a vector of current endogenous variables.

Y_{t-1} is a vector of endogenous variables lagged one time period.

X_t is a vector of current exogenous variables.

X_{t-1} is a vector of exogenous variables lagged one time period.

u_t^* is a vector of derived reduced form disturbances where each element denotes the disturbance associated with an individual reduced form equation.

A is the coefficient matrix of the lagged endogenous variables.

B is the coefficient matrix of the current exogenous variables.

C is the coefficient matrix of the lagged exogenous variables.

There are several characteristics of the analytically derived reduced forms specified in (19.19) worth noting. First, both lagged endogenous and exogenous variables are included in the equations. Thus the structural equations have to be formulated in such a manner so as to include these lagged variables. Second, each model need not contain all endogenous and exogenous variables lagged. Some elements of A may be zero which indicates that certain lagged endogenous variables are not included in the system. Similarly, some elements of C may be zero indicating that certain lagged exogenous variables are not included in the system. Of course, just which lagged variables will be included and which will be excluded depends on the problem being analyzed and the phenomena being modelled. Finally, if interest is in impact multiplier and tracing the effects of changes in exogenous variables

---

[2/] For a detailed discussion see Theil, H. and Boot, J. C. G. "The Final Form of Econometric Equation Systems", Review of the International Statistical Institute, Vol. 30: 2, 1962, pp. 136-152.

through time then A and C must not be null matrices.

Impact Multipliers

Impact multipliers measure the immediate net effect of changes in exogenous variables on endogeneous variables. In other words the net effects of changes in exogenous variables in time t on endogenous variables in time t. In (19.19) the elements of B are the impact multipliers.

Interim Multipliers

Interim multipliers give the net effects of changes in exogenous variables in time t on endogenous variables in time t' where t'>t. The interim multipliers may be expressed as the following derivative:

$$(19.20) \quad \frac{\partial Y_{t'}}{\partial X_t} \qquad t' > t$$

Let us consider the net effect of changes in exogenous variables on endogenous variables one year later. Thus $t' - t = s = 1$. We have,

$$(19.21) \quad Y_{t-1} = AY_{t-2} + BX_{t-1} + CX_{t-2} + u^*_{t-1}$$

and substituting in (19.19) yields:

$$(19.22) \quad Y_t = A(AY_{t-2} + BX_{t-1} + CX_{t-2} + u^*_{t-1}) \; BX_t + CX_{t-1} + u^*_t$$

$$= A^2 Y_{t-2} + BX_t + (AB + C) X_{t-1} + ACX_{t-2} + (u^*_t + Au^*_{t-1})$$

Thus, the influence of changes in exogenous variables on endogenous variables one year later is given by:

$$(19.23) \quad (AB + C)$$

To find the influence of changes in exogenous variables on endogeneous variables two years later we consider $t' - t = s = 2$. We have

$$(19.24) \quad Y_{t-2} + AY_{t-3} = BX_{t-2} + CX_{t-3} + u^*_{t-2}$$

Substituting into (19.22) yields:

$$(19.25)\quad Y_t = A^2(AY_{t-3}+BX_{t-2}+CX_{t-3}+u^*_{t-2}) + BX_t + (AB+C)X_{t-1}+ACX_{t-2} +(u^*_t+Au^*_{t-1})$$

$$= A^3Y_{t-3}+A^2BX_{t-2}+A^2CX_{t-3}+A^2u^*_{t-2}+BX_t +(AB+C)X_{t-1}+ACX_{t-2}+(u^*_t+Au^*_{t-1})$$

$$= A^3 Y_{t-3} + BX_t + (AB + C) X_{t-1} + A(AB+C) X_{t-2} + A^2CX_{t-3}$$

$$+ (u^*_t + Au^*_{t-1} + A^2u^*_{t-2})$$

Thus the influenee of changes in exogenous variables on endogenous variables two years later is given by,

$$(19.26)\quad A(AB + C)$$

This process could be continued for any length of lag. In general let $t' - t = s$ and

$$(19.27)\quad D^s = A^{s-1}(AB + C)$$

D^s is a measure of the net effects of changes in exogenous variables on endogenous variables s time periods later. Again, note that some lagged endogneous variales have to be included in model or it will not be possible to assess the interim multipliers.

Total Multipliers

Total multipliers refer to the accumulated net effect of changes in exogenous variables on endogenous variables as the time lapse approaches infinity. That is $t' - t = s \to \infty$. In other words total multipliers measure the accumulated affects of all possible interim multipliers or they are measuring the net affects of changes in exogenous variables from the time the changes occur until the effects are dissipated as the time elapsed approaches infinity. The total multipliers are the infinite sum:

$$(19.28)\quad \sum_{t'=t}^{\infty} \frac{\partial Y(t')}{\partial X(t)}$$

provided the sum converges to a finite limit.

In terms of the coefficient matrices of the analytically derived reduced forms we have been considering (19.28) are elements of:

$$(19.29) \quad B + \sum_{s=1}^{\infty} A^{s-1} (AB + C) = B + (I - A)^{-1} (C + AB)$$

Given the analytically derived reduced form equations in (19.19), the total accumulated effects of changes in exogenous variables can easily be assessed using (19.29).

Summary

Impact multipliers, interim multipliers, and total multipliers provide the type of information often sought by policy makers. These multipliers are obtained from analytically derived reduced forms which are derived from estimated structural equations. Thus, the usefulness of structural estimation may not be to obtain estimated structural parameters per se but to obtain various types of impact multipliers which may be helpful in policy evaluation and decision making.

CHAPTER 20
Seemingly Unrelated Regressions

Introduction

Seemingly unrelated regressions (SUR) estimation is concerned with efficiently estimating the coefficients of a set of linear regression equations that are related through their disturbances. The method proposed by Zellner[1/] uses Aitken's generalized least squares (GLS) procedures to jointly estimate the coefficients of a set of linear regression equations. Because of this joint estimation the method is sometimes referred to as joint generalized least squares (JGLS).

SUR or JGLS estimation deals with estimating a combination of several linear regression equations which appear unrelated but which in fact are related because the equations contain some predetermined variables that are common to each equation and the disturbances of the equations are correlated. Under the condition that the disturbances of the different regression equations are correlated the use of GLS in a second stage yields estimators that are asymptotically more efficient than estimators obtained by ordinary least squares (OLS) applied to each equation separately. Furthermore, the SUR estimator's variances are not far different from their asymptotic variances for a wide range of correlation coefficient values between disturbances of the regression equations of the set. Also, for sample sizes of 20 or more the density function of the SUR estimators is very close to a normal density. If the disturbances among the regression equations of the set are not correlated the SUR estimators are identical to OLS estimators and there are no gains in efficiency possible.

[1/] Detailed information about estimation procedures and some finite sample properties are given in two articles by Arnold Zellner The articles are: Arnold Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias" Journal of the American Statistical Association, Vol. 57, No. 298, June 1962, pp. 348-368 and Arnold Zellner "Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results", Journal of the American Statistical Association, Vol. 58, No. 304, December 1963, pp. 977-992.

In this chapter attention will first be devoted to a couple illustrative examples where SUR estimators are applicable. Next focus will be on the procedure for estimating SUR estimators. Then the estimation of the variance - covariance matrix of the disturbances will be discussed. This will be followed by an application example of SUR estimation and the chapter will close with a comment of gains in efficiency by SUR estimation.

Application of SUR Estimation

Consider a situation where we are dealing with a set of M linear multiple regression relationships for which all the assumptions underlying ordinary least squares are fulfilled for each of the equations individually. In particular the flow of causality is uni-directional from the predetermined variables to the single endogenous variable in each equation. Let us assume that all the assumptions concerning the disturbance of each equation are fulfilled in order to obtain desireable properties from OLS estimation. Although there may be some predetermined variables that are common to each equation, let there be at least one predetermined variable in each equation that does not appear in any of the other equations of the set. There of course may be more than one such predetermined variable but there must be at least one such variable. Furthermore, the number of predetermined variables per equation may vary. The disturbance term of each equation represents the effects of all variables not explicitly included in the equation. Let us assume that the effects represented by the disturbance term of each equation is the combined effect of a set of excluded variables that are common to all equations and thus lead to a non-zero correlation among the disturbances of the equations. Finally, we will be considering an adequate sample size of n observations for each variable, endogenous and predeterminad, in each equation. Under these conditions SUR estimation will yield more efficient estimators than will ordinary least squares.

As the first illustration let us consider the demand by individual farmers for the fertilizer nutrients nitrogen (N),phosphoric acid (P), and potosh (K). The model hypothesized to accounted for the quantity demanded of each of the three nutrients entails the prices of the three nutrients, prices of the relevant outputs for which these fertilizers are used, selected attributes of the farms and of the farm operators hypothesized to influence the quantity demanded of each of the nutrients. The economic unit, in this case farmers, are assumed not to influence through their behavior the market prices of N, P and K and the market prices of the relevant outputs. Also, the selected attributes of the farm operators and the farms themselves are assumed not to be affected by the behavior of the farm operator or farmer in his purchasing of N, P, and K. Under these specifications the causality appears to be uni-directional and the formulation of single equations to be estimated by ordinary least squares methods seems to be quite appropriate. However, let us assume that each individual farm operator purchased all three of these fertilizers in the same market, however, you wish to define market and thus the disturbance represents the influence of variables not included in the equations that are common to all three nutrient equations and furthermore they are variables which pertain to the same market. Thus one could suspect that the disturbances for the three fertilizers would be correlated and that SUR estimation would result in some gains in efficiency in terms of small variances of at least some and possibly all of the estimated coefficients.

As a second illustrative example let us consider the simple investment equations considered by Zellner in illustrating the application of SUR estimation.[2/] The investment functions involve a firm's gross investment being dependent on the firm's beginning-of-year capital stock, and the value of its outstanding shares at the beginning of the year. To estimate these micro-investment functions two firms were selected, General Electric and Westinghouse.

[2/] Zellner, op.cit., 1962, pp. 357-362.

It was assumed that the disturbance term represented a set of forces common to both of these firms and that the disturbance would likely be correlated. And in fact they were and the SUR estimators were more efficient. The estimates of the variances of the SUR estimated coefficients were about 20 per cent smaller than the estimated variances of the coefficients estimated equation-by-equation using ordinary least squares methods.

Estimation Procedure

Suppose we are interested in estimating the following set of two equations:

$$Y_1 = X_1\beta_1 + u_1 \tag{20.1}$$
$$Y_2 = X_2\beta_2 + u_2$$

where

Y_1 is a nx1 vector of observations of the first endogenous variable.

Y_2 is a nx1 vector of observations of the second endogenous variable.

X_1 is a nxk_1 matrix of observations of the predetermined variables in the first equation.

X_2 is a nxk_2 matrix of observations of the predetermined variables in the second equation.

β_1 is the k_1x1 vector of unknown parameters of the first equation

β_2 is the k_2x1 vector of unknown parameters of the second equation.

u_1 is the nx1 vector of disturbances of the first equation.

u_2 is the nx1 vector of disturbances of the second equation.

We are assuming that:

$$E(u_1) = E(u_2) = 0 \tag{20.2}$$

and

$$E(u_1^2) = \sigma_{11} I \tag{20.3}$$

and

$$E(u_2)^2 = \sigma_{22} I \tag{20.4}$$

now satisfying least squares assumptions the least squares estimators of β_1 and β_2 are:

(20.5) $\hat{\beta}_1 = (X_1' X_1)^{-1} X_1' Y_1$

and

(20.6) $\hat{\beta}_2 = (X_2' X_2)^{-1} X_2' Y_2$

The estimators in (20.5) and (20.6) are the results of applying ordinary least squeres to each of the two equations separately.

Now let us combine the two equations in (20.1) by rewriting them as:

(20.7) $$Y_1 = [X_1 : 0] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + u_1$$

and

(20.8) $$Y_2 = [0 : X_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + u_2$$

Then define

$$Y = \begin{bmatrix} Y_1 \\ \hline Y_2 \end{bmatrix}, \qquad X = \left[\begin{array}{cc} X_1 & 0 \\ \hline 0 & X_2 \end{array}\right], \qquad u = \begin{bmatrix} u_1 \\ \hline u_2 \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_1 \\ \hline \beta_2 \end{bmatrix}$$

and rewrite the combined set as:

(20.9) $Y = X\beta + u$

If u_1 and u_2 are uncorrelated we have:

(20.10) $\beta = (X'X)^{-1}X'Y$

If u_1 and u_2 are correlated the SUR set of estimators are an efficient set of estimators and are obtained in the following manner:

(20.11) $\hat{\beta}^* = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$

where

(20.12) $$\beta^* = \begin{bmatrix} \hat{\beta}_1^* \\ \hline \hat{\beta}_2^* \end{bmatrix}$$

and

(20.13) $$\Sigma = \begin{bmatrix} \sigma_{11} I & \sigma_{12} I \\ \sigma_{21} I & \sigma_{22} I \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \otimes I$$

Σ is the variance-covariance matrix of the disturbances of the equations. The latter expression of (20.13) is the Kronecker product expression of the variance-covariance matrix of the u's.

Let the inverse of the variance-covariance matrix of the disturbances be:

$$(20.14)\qquad \Sigma^{-1} = \begin{bmatrix} \sigma^{11} I & \sigma^{12} I \\ \sigma^{21} I & \sigma^{22} I \end{bmatrix} = \begin{bmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{21} & \sigma^{22} \end{bmatrix} \otimes I$$

then

$$(20.15)\qquad \hat{\beta}^* = \begin{bmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2^* \end{bmatrix} = \begin{bmatrix} \sigma^{11} X_1'X_1 & \sigma^{12} X_1'X_2 \\ \sigma^{21} X_2'X_1 & \sigma^{22} X_2'X_2 \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_{j=1}^{2} \sigma^{1j} X_1'Y_j \\ \sum_{j=1}^{2} \sigma^{2j} X_2'Y_j \end{bmatrix}$$

The variance-covariance matrix of the estimator $\hat{\beta}^*$ is:

$$(20.16)\qquad \text{Var}\ (\hat{\beta}^*) = (X' \Sigma^{-1} X)^{-1}$$

or

$$(20.17)\qquad \text{Var}\ (\beta^*) = \begin{bmatrix} \sigma^{11} X_1'X_1 & \sigma^{12} X_1'X_2 \\ \sigma^{21} X_2'X_1 & \sigma^{22} X_2'X_2 \end{bmatrix}^{-1} .$$

Thr procedure can be generalized to and M > 2 equations by letting (20.9) represent the following set of equations:

$$(20.18)\qquad \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 & 0 & - & - & 0 \\ 0 & X_2 & 0 & 0 & - & - & 0 \\ 0 & 00 & X_3 & 0 & - & - & 0 \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ 0 & 0 & - & - & - & - & X_M \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \\ \\ \\ \beta_M^* \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \\ \\ \\ u_M \end{bmatrix}$$

and the variance-covariance matrix of the disturbances of the equations is:

$$(20.19)\qquad \Sigma = \begin{bmatrix} \sigma_{11} I & \sigma_{12} I & - & - & - & \sigma_{1M} I \\ \sigma_{21} I & \sigma_{22} I & - & - & - & \sigma_{2M} I \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ \sigma_{M_1} I & \sigma_{M_2} I & - & - & - & \sigma_{MM} I \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & - & - & - & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & - & - & - & \sigma_{2M} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \sigma_{M1} & \sigma_{M2} & - & - & - & \sigma_{MM} \end{bmatrix} \otimes I$$

and

$$(20.20) \quad \Sigma^{-1} = \begin{bmatrix} \sigma_{11}I & \sigma_{12}I & - & - & - & \sigma_{1M}I \\ \sigma_{21}I & \sigma_{22}I & - & - & - & \sigma_{2M}I \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \sigma_{M1}I & \sigma_{M2}I & - & - & - & \sigma_{MM}I \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \sigma^{11}I & \sigma^{12}I & - & - & - & \sigma^{1M}I \\ \sigma^{21}I & \sigma^{22}I & - & - & - & \sigma^{2M}I \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \sigma^{M1}I & \sigma^{M2}I & - & - & - & \sigma^{MM}I \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^{11} & \sigma^{12} & - & - & - & \sigma^{1M} \\ \sigma^{21} & \sigma^{22} & - & - & - & \sigma^{2M} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \sigma^{M1} & \sigma^{M2} & - & - & - & \sigma^{MM} \end{bmatrix} \otimes I$$

The estimators for the M equations are:

$$(20.21)\quad \hat{\beta}^* = \begin{bmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \cdot \\ \cdot \\ \cdot \\ \hat{\beta}_M^* \end{bmatrix} = \begin{bmatrix} \sigma^{11}X_1'X_1 & \sigma^{12}X_1'X_2 & \cdots & \sigma^{1M}X_1'X_M \\ \sigma^{21}X_2'X_1 & \sigma^{22}X_2'X_2 & \cdots & \sigma^{2M}X_2'X_M \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma^{M1}X_M'X_1 & \cdots & & \sigma^{MM}X_M'X_M \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_{j=1}^{M} \sigma^{1j} X_1'Y_j \\ \sum_{j=1}^{M} \sigma^{2j} X_2'Y_j \\ \cdot \\ \cdot \\ \cdot \\ \sum_{j=1}^{M} \sigma^{Mj} X_M'Y_j \end{bmatrix}$$

and

$$(20.22)\quad \text{var}\,(\hat{\beta}^*) = \begin{bmatrix} \sigma^{11}X_1'X_1 & \sigma^{12}X_1'X_2 & \cdots & \sigma^{1M}X_1'X_M \\ \sigma^{21}X_2'X_1 & \sigma^{22}X_2'X_2 & \cdots & \sigma^{2M}X_2'X_M \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma^{M1}X_M'X_1 & \cdots & & \sigma^{MM}X_M'X_M \end{bmatrix}^{-1}$$

The estimators presented in (20.11), (20.12), (20.15) and (20.21) consider Σ to be known. Most frequently this is not the case and Σ has to be estimated from the residuals of the equations estimated by ordinary least squares. The estimation of the variance-covariance matrix is discussed in the next section.

Estimating the Variance-Covariance Matrix

Since usually Σ is unknown it needs to be estimated from the residuals of the equations estimated separately by ordinary least squares methods. The least squares estimators are given in (20.10). However, instead of letting these pertain to only two, let (20.10) pertain to M equations. The estimated residuals for each equation may be obtained as:

$$(20.23)\quad (Y - X\hat{B}) = \hat{U}$$

which may be expressed in more detail as:

$$(20.24)\qquad \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_M \end{bmatrix} - \begin{bmatrix} X_1 & 0 & 0 & \cdots & 0 \\ 0 & X_2 & 0 & \cdots & 0 \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ 0 & \cdots & \cdots & \cdots & X_M \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\beta}_M \end{bmatrix} = \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{u}_M \end{bmatrix}$$

The variance may now be estimated as:

$$(20.25)\qquad s_{jj} = \frac{u_j' u_j}{n - k_j}$$

and the estimated covariances between the disturbances of the $i^{\underline{th}}$ and $j^{\underline{th}}$ equation based on the least squares residuals between the $i^{\underline{th}}$ and $j^{\underline{th}}$ equations are:

$$(20.26)\qquad s_{ij} = \frac{u_i' u_j}{\sqrt{\left[n - k_i\right]\left[n - k_j\right]}}$$

where k_i and k_j are the degrees of freedom "lost" in each of the respective equations. Now the estimated variance-covariance matrix of the disturbances is:

$$(20.27)\qquad \tilde{\Sigma} = \begin{bmatrix} s_{11}I & s_{12}I & \cdots & s_{1M}I \\ s_{21}I & s_{22}I & \cdots & s_{2M}I \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ s_{m1}I & s_{M2}I & \cdots & s_{MM}I \end{bmatrix}$$

$$= \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1M} \\ s_{21} & s_{22} & \cdots & s_{2M} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ s_{M1} & s_{M2} & \cdots & s_{MM} \end{bmatrix} \otimes I$$

The inverse of the estimated variance-covariance matrix of disturbances is:

$$(20.28)\quad \tilde{\Sigma}^{-1} = \begin{bmatrix} s^{11} & s^{12} & \cdots & s^{1M} \\ s^{21} & s^{22} & \cdots & s^{2M} \\ \vdots & & & \vdots \\ s^{M1} & s^{M2} & \cdots & s^{MM} \end{bmatrix} \otimes I$$

where

$$(20.29)\quad \left[s^{ij}\right] = \left[s_{ij}\right]^{-1}$$

The estimators using the estimated variance-covariance matrix are:

$$(20.30)\quad \tilde{\beta}^{*} = \begin{bmatrix} \tilde{\beta}_1^{*} \\ \tilde{\beta}_2^{*} \\ \vdots \\ \tilde{\beta}_M^{*} \end{bmatrix} \begin{bmatrix} s^{11}X_1'X_1 & s^{12}X_1'X_2 & \cdots & s^{1M}X_1'X_M \\ s^{21}X_2'X_1 & s^{22}X_2'X_2 & \cdots & s^{2M}X_2'X_M \\ & & \vdots & \\ s^{M1}X_M'X_1 & & \cdots & s^{MM}X_M'X_M \end{bmatrix} \times \begin{bmatrix} \sum_{j=1}^{M} s^{1j}X_1'Y_j \\ \sum_{j=1}^{M} s^{2j}X_2'Y_j \\ \\ \sum_{j=1}^{M} s^{Mj}X_M'Y_j \end{bmatrix}$$

The estimated variance-covariance matrix of the estimators presented in (20.30) is:

$$(20.31)\quad \text{var}\,(\tilde{\beta}^{*}) = \begin{bmatrix} s^{11}X_1'X_1 & s^{12}X_1'X_2 & \cdots & s^{1M}X_1'X_M \\ s^{21}X_2'X_1 & s^{22}X_2'X_2 & \cdots & s^{2M}X_2'X_M \\ & & \vdots & \\ s^{M1}X_M'X_1 & s^{M2}X_M'X_2 & \cdots & s^{MM}X_M'X_M \end{bmatrix}^{-1} + O(n^{-1})$$

where $O(n^{-1})$ denotes a bias term of higher order of smallness than n^{-1} where n is the sample size. From a practical viewpoint the $O(n^{-1})$ term is usually disregarded.

An Application

To illustrate the SUR estimation procedure the example given by Zellner (1962) will be highlighted here. The model considered is:

$$(20.32) \quad I_t = \alpha_0 + \alpha_1 C_{t-1} + \alpha_2 F_{t-1} + u_t$$

where

I_t is a firm's current gross investment.

C_{t-1} is the firm's beginning-of-year capital stock.

F_{t-1} is the value of the firm's outstanding shares at the beginning of the year.

The two firms considered were General Electric and Westinghouse. Annual data for the period 1935-1954 were used to estimate (20.32) by single equation ordinary least squares and the two-stage Aitken's procedure which yields SUR estimators.

Following Zellner, for convenience, we relabel the variables as follows:

Firm	I(t)	C(t-1)	F(t-1)	1
General Electric	$Y_{1(t)}$	$X_{11(t)}$	$X_{12(t)}$	$X_{13(t)}$
Westinghouse	$Y_{2(t)}$	$X_{21(t)}$	$X_{22(t)}$	$X_{23(t)}$

The equation system to be estimated is:

$$(20.33) \quad \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

where

$$X_1 = \begin{bmatrix} X_{11(t)} & X_{12(t)} & X_{13(t)} \end{bmatrix} \quad X_2 = \begin{bmatrix} X_{21(t)} & X_{22(t)} & X_{23(t)} \end{bmatrix}$$

$$\beta_1' = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{10} \end{bmatrix} \quad \text{and} \quad \beta_2' = \begin{bmatrix} \beta_{21} & \beta_{22} & \beta_{20} \end{bmatrix}$$

Zellner first computed the single-equation ordinary least squares estimates in the usual way. Then he obtained the variance-covariance matrix of the estimated disturbances using (20.25) and (20.26) to obtain the variances and covariances and (20.27) to get the variance-covariance matrix. He obtained the inverse as in (20.28) and then utilized (20.30) and (20.31) to obtain the two-stage Aitken's (SUR) estimators and their respective variances. These estimators and their variances and the single-equation ordinary least squares estimators and their respective variances are presented in Table 1.

Table 1.

Results of Two-Stage Aitken and Single-Equation Least-Squares Estimation of Micro-Investment Functions

Micro-Unit	Coefficient of	Two-Stage Aitken Method		Single-Equation Least Squares	
		Coefficient Estimate	Variance of Coefficient Estimator	Coefficient Estimate	Variance of Coefficient Estimator
General Electric	C-1	.1326	.0006006	.1517	.0006605
	F-1	.0421	.0001885	.0266	.0002423
	1	-32.4807	789.6	-9.9563	984.1
Westing-house	C-1	.0459	.002691	.0924	.003147
	F-1	.0611	.0001972	.0529	.0002468
	1	-2.0113	54.21	-.5094	64.24

It is seen from this table that the application of the estimation procedure described above has resulted in approximately a 20% reduction in the estimated coefficient estimator variances, a significant reduction, compared with those of single-equation least squares. Notice also that the coefficient estimates yielded by the two methods differ. This is expected since each method is minimizing a different quadratic form, and since one method is more efficient than the other, the estimates usually differ.

The logic of using this method to the investment equations of both firms. General Electric and Westinghouse is that these firms operate in the same branch of industry, it is conceivable that there are common factors that affect the investment decisions of both firms quite apart from the common

development of the two explanatory variables, C-1 and F-1 of each firm. Since these factors are neglected by the model, they have to be analyzed in terms of disturbances. The disturbances of the General Electric investment equation are to be expected to be correlated with those of the Westinghouse equation, if there are indeed such common factors. An indication of this is that the 20 LS residuals of General Electric and the 20 LS residuals of Westinghouse have a correlation coefficient of .73.

Gains in Efficiency

The gain in efficiency yielded by Zellner's SUR estimators over ordinary least squares estimators increases directly with correlation between the disturbances from different equations and inversely with the correlation between the different sets of predetermined variables.

Even if the true correlation between equation disturbances is zero, the sample least squares residuals may yield non-negligable covariances and SUR estimators might be computed mistakenly. For finite samples the result will be estimates with somewhat greater standard errors then those of the ordinary least squares coefficients. This will even be true for very small correlations among the disturbances but as these correlations increase the efficiency of the SUR estimators over the ordinary least squares estimators rises substantially.

To illustrate the effect on efficiency of using an estimated disturbance variance-covariance matrix Zellner computed the exact second moment matrix of the SUR estimators[3/] which is:

$$(20.34)\quad E(\tilde{\beta}_1^* - \beta_1^*)(\tilde{\beta}_1^* - \beta_1)' = (X_1'X)^{-1}\,\sigma_{11}(1-\rho^2)\left[1 + \frac{.5649}{d-2}\,\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(d/2\right)}\right]$$

[3/] This is based on a two equation model.

where $d = n - k_1 - k_2$

n is sample size.

k_1 is the degrees of freedom "lost" from equation one.

k_2 is the degrees of freedom "lost" from equation two.

Γ is a gamma function.

ρ is the correlation coefficient between disturbances.

The value of $(1 - \rho^2)$ decreases as ρ increases in absolute terms. If

$$(20.35) \quad (1 - \rho^2) \left[1 + \frac{.5649}{d-2} \quad \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(d/2\right)} \right]$$

is equal to one, the result is the ordinary least squares variance-covariance matrix of $\hat{\beta}_1$ i.e. $(X_1'X)^{-1} \sigma_{11}$, based on a known variance-covariance matrix of disturbances. For an estimated variance-covariance matrix of disturbances among equations Zellner tabulated the value of (20.35) for wide range of values of d and ρ. The results basically showed the substantial gains in efficiency occurred for large degrees of freedom and high values of the correlation coefficient among the disturbances.

CHAPTER 21

Recursive Models

Introduction

A central concern of econometrics is the building of models which are simplified representations of the real world. One of the long standing controversies of the art of model building is whether economic magnitudes should be specified as recursively or simultaneously determined. These alternative simplifications of the model of the real world have important implications from both the theoretical and empirical viewpoints. Herman Wold has been an almost continuous crusader for his view of the world as essentially recursive, since his early publications on the topic.[1/]

In this chapter the foundations of recursive models will be outlined and the controversy of recursive chains versus simultaneous equation models will be discussed. In the next section the theoretical basis for recursive chains as economic model representations will be discussed. This will be followed by a discussion of some empirical questions related to the use of recursive chain models. Then some comparisons are made between recursive models and simultaneous equation models. In the final section the strong points of recursive models and simultaneous equation models are summarized.

1/ For details of the development of causal chain concepts and recursive models see the following:

1. Wold H. and Jureen L., Demand Analysis, Wiley and Sons, 1953.
2. Wold H. "Causality and Econometrics", Econometrica, Vol. 22, No. 2, April, 1954.
3. Strotz, R. and Wold H., "Recursive Versus Non-Recursive Systems: An Attempt to Synthesis", Econometrica, Vol. 28, No. 2, April, 1960.
4. Wold, H., "A Generalization of Causal Chain Models", Econometrica, Vol. 28, No. 2, April 1960.

Foundations for Recursive Systems

Econometricians use an economic model as a simplified representation of the real world. The basis for a recursive model is the hypothesis that behavior in the real world is recursive. This implies that behavior is unidirectional and that regardless of the length of time period being considered a particular pattern of time elapse occurs between the changes in variables that act as stimuli and changes in variables that act as responses to those stimuli. Recursive models consist of systems of equations containing both endogenous and exogenous variables with one causal relationship for each endogenous variable. Given exogenous variables and initial values of endogenous variables, the development of endogenous variables in successive time periods may be calculated recursively.

When considering recursive models it is important to distinguish between deterministic economic models and stochastic econometric models. Proponents of recursive models have advocated their use in both situations but have acknowledged that the recursive approach is especially suited to the theoretical representation of the economic model. Bentzel and Hansen[2/] concerned themselves with whether or not there is in pure economic theory, an a priori case for or against the use of recursive or simultaneous equation models. They distinguish between "basic" models which are recursive and "derived" models (more simplified ones) which may not be recursive.

The rationale underlying recursive models has been presented in a framework of a theoretical definition of causality. Scientists use a controlled experiment to establish a causal relationship,they specify X as the cause and Y as the effect and hold all other factors constant to see how variations in the cause X's result in variations in effect Y. Such experiments can be

---

2/ Bentzel R. and Hansen B. "One Recursiveness and Interdependency in Economic Models", Review of Economic Studies, Vol. 22, 1954-55.

replicated and verified or reputed as working hypotheses. Economists make their experiments without control, and yet attempt to specify the direction of causality. They say the demand relation $d = D(P)$ is causal but that the inverse $p = D^{-1}(d)$ is not. The relation is thus asymetric since causality runs one way only. The problem is that since variations in cause, X do not occur in a constant environment, the very statement of the hypothesis "X causes Y" may be questionable. Thus the specification problem needs to be solved before the application of statistical tools to an empirical model.

Bentzel and Hansen refer to a basic model as one in which complete account is taken of each economic good or subject. These economic subjects are the building blocks of theoretical economic models· they have expectations which form the basis of plans of action, then actions; finally there is accounting of the outcomes of the interaction of economic subjects with one another, The elements of expectations, plans, actions and accounting follow one another in recursive manner.

At the point of division between two time periods economic subjects make new plans (based on expectations which are influenced by events in the preceding time period). These plans may be conditioned or unconditioned· for example, purchasing plans conditioned by price quotations. These plans are interdependent, but based on a set of recursive (conditioned) demand equations. Through the period conditioned plans become less conditioned in a chain of reactions. The purchaser (seller) carries out a plan that eventually makes the seller's (purchaser's) conditioned plan become unconditioned. All human actions take time so this procedure must be recursive rather than simultaneous.

The above argument seems to be the most basic underlying one in the literature for a view of the whole world as being recursive. It merits careful consideration, keeping in mind that it refers only to the theoretical economic model, and then only to the basic economic subject. An economic subject exists only in what might be called the basic model, and it is limited to a

single decision making and plan formulating unit. A firm with separate purchasing sales divisions, for example, may be seen as making simultaneous decisions. However, some would argue that it is the economic unit that has divided into two behaving units, each taking actions recursively. Bentzel and Hansen have thus implicitly defined recursively at the micro level, as the successive unconditioning of previously conditioned plans of action. Looking at the whole firm is aggregating over more than one economic subject.

An important point in the above argument for recursivity in basic models, is the definition of time periods. In theory, the time interval in which plans go from conditioned to unconditioned and materialize as actions, may be defined to be as short or long as is necessary to assume recursivity of behavior of the economic subject. In a sense this is assuming away the controversy of recursive versus simultaneous models., but, this has important implications for the use of derived econometric as distinct from basic economic models.

Recursivity may not apply to derived models which are simplified and omit some variables and relations from the basic model. This is especially so for econometric models which always use ex poste variables and lack knowledge of empirical values of ex anti variables. Several specific reasons are given for interdependency in such derived models. Most importantly, these models aggregate over time periods in which plans materialize into actions, and over separate economic subjects. For example, aggregation over households makes them interdependent and their combined plans can be specified in the derived model as simultaneous. This leads to specification of variables as simultaneously determined when they are in fact recursively determiend over a different time interval. Another related point is the use of a static framework with equilibrium conditions to simplify the econometric model. This static view results from a lack of empirical information on the expectations and plans of economic subjects. Interdependency in dynamic models results

from conceiving the differences between ex ante and ex poste variables as random variables.

If it is accepted that the basic economic models is a recursive one, then the controversy of simultaneous versus recursive models revolves around whether the derived econometric models is simultaneous or recursive. Put another way, do the simplification and modifications contained in the econometric model require simultaneous or recursive specifications?

Empirical Recursive Models

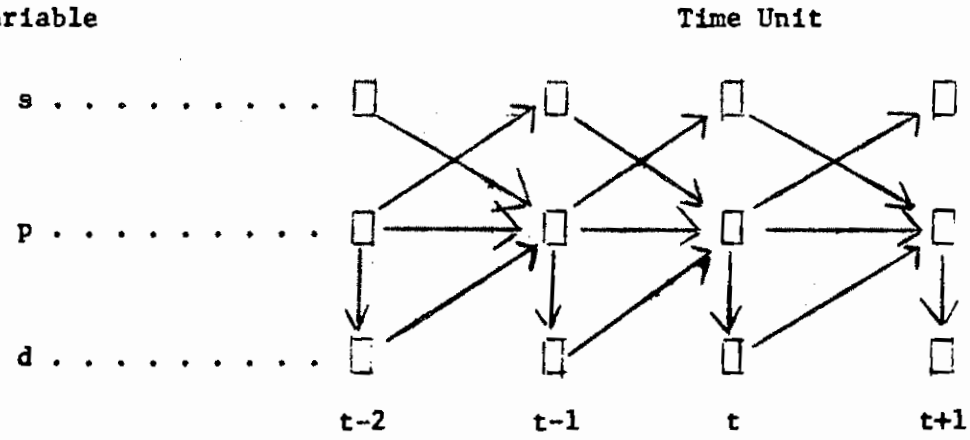
Wold and others take the view that empirical derived models should be specified as recursive. In a statistical analysis of demand and supply Wold uses the following model:

(21.1) $d_t = D(pt)$

(21.2) $s_t = S(p_t - 1)$

(21.3) $p_t = p_{t-1} + \gamma (d_{t-1} - s_{t-1})$

Equation (21.1) is a causal hypothesis of consumers' demand as a function of current price. Equation (21.2) shows supply a causally related to price in the previous production period. Equation (21.3) indicates price adjustments through time and the causal mechanism behind price movements. Given initial values of p_0 and s_0 we can recursively calculate the development of d_t, s_t, and p_t as represented in the follow arrow diagram.



As the arrow directions indicate, values of the variables can be determined recursively through time.

These considerations lead us to the following type of recursive system:

(21.4) $Y_{1t} = 0 + 0 + a_{11}Z_{1t} + u_{1t}$

(21.5) $Y_{2t} = b_{21}Y_{1t} + 0 + a_{22}Z_{2t} + u_{2t}$

(21.6) $Y_{3t} = b_{31}Y_{1t} + b_{32}Y_{2t} + a_{33}Z_{3t} + u_{3t}$

where the Y's are endogenous variables, the Z's are lagged endogenous and exogenous variables and u's are random disturbances. The a's are coefficients of predetermined variables while the b's are coefficients of the endogenous variables. Unilateral causal dependence is hypothesized so in the second and third equations of the system, the Y's on the right hand side are predetermined endogenous variables. The matrix of coefficients of the endogenous variables is triangular but all elements of the triangular matrix need not be different from zero. In other words, all of the endogenous variables in equations above the one being considered do not need to appear in the equation being considered. Specifically, both Y_{1t} and Y_{2t} need not appear in (21.6).

Z_{1t}, Z_{2t}, and Z_{3t} may have some common predetermined variables but each must contain one or more that are not conditioned in the other equations. If this were not the case, then a single endogenous variables could be substituted for an entire set of predetermined variables. For example suppose Z_{1t} and Z_{2t} were identical. Then (21.5) really would be:

$$(21.7) \quad Y_{2t} = b_{21}Y_{1t} + a_{22}Y_{1t} + u_{2t}$$

$$= (b_{21} + a_{22}) Y_{1t} + u_{2t}$$

or

$$(21.8) \quad Y_{2t} = b_{21}(a_{11}Z_{1t}) + a_{22}Z_{2t} + u_{2t}$$

$$= b_{21}a_{11}Z_{1t} + a_{22}Z_{2t} + u_{2t}$$

And if Z_{1t} and Z_{2t} are identical then,

$$(21.9) \quad Y_{2t} = (b_{21}a_{11} + a_{22}) Z_{2t} + u_{2t}$$

$$= (b_{21}a_{11} + a_{22}) Z_{1t} + u_{2t}$$

Based on (2.19) one would have to conclude that Y_{2t} is a function of only predetermined variables and that the equation is not linked to any other equation of the system.

Assumptions about the disturbances used in recursive empirical models are important with respect to the estimation procedure. Both recursive and

simultaneous equation systems models assume predetermined variables (Z's) and disturbances (u's) are uncorrelated. Recursive models require in addition that u_i is uncorrelated with Y_{1t}, Y_{2t}, - - - $Y_{i-1,t}$ and that at least some of the Z's from successive equations are independent. In simultaneous equation models the Y_{it} endogenous variables are correlated with the disturbances of a recursive system are met the variance-covariance matrix of the disturbance terms is diagonal. If the model is specified correctly, (i.e., if the true structure is recursive as hypothesized), then these assumptions will hold and ordinary least squares (OLS) is appropriate for estimation. This is especially important since factors excluded from the derived model enter through the disturbances. Since it is possible that omitted variables may be correlated with included variables zero correlations may complicate the model.

The use of OLS in estimation of recursive models under the above assumptions, and all the "usual" assumptions for OLS results in BLUE estimates of the parameters of the causal system. If OLS is used on simultaneous systems, on the other hand, such desirable properties are not present. Parameter estimates are biased inconsistent and inefficient. Regression analysis was originally developed for use in controlled experiments where unilateral causal dependence could be stated. In non-controllable economic data the assumptions of OLS may be questioned, and specification of the model being estimated as simultaneous, would result in biased inconsistent and inefficient parameter estimates.

Whether the derived econometric model is to be used for estimation of structural parameters, or for forecasting and prediction, affects the comparison of empirical estimation. In the latter case, we do not need to know structural parameters, and all that is required is that the unknown structure remains unchanged.

Cournot and Walras:

The models were concerned with static Equilibrium, and the emphasis was on stability conditions. Cournot drew reaction functions and showed how duopolists adjusted prices in a recursive manner, one following the other with a decision based on predetermined variables. Walrasian equilibrium was based on predetermined variables. Walrasian equilibrium was based on

(21.10) $d = D(p)$

(21.11) $s = S(p)$

and equilibrium occurred where

(21.12) $D(p^*) = S(p^*)$

which may be shown as follows:

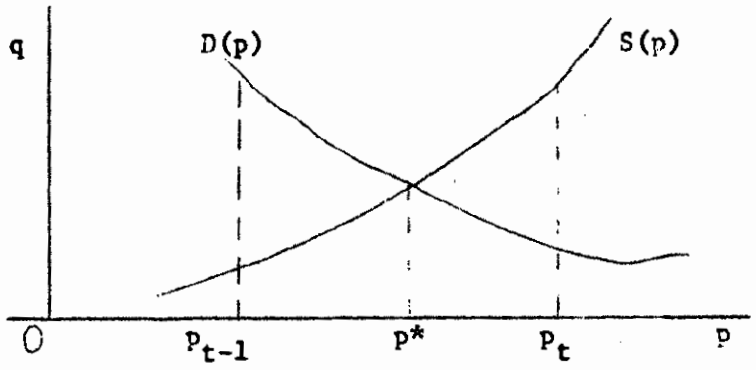


FIGURE 21.1

Cobweb Models

Let us consider the following model:

(21.13) D: $p_t = -aq_t + u_t$

(21.14) S: $q_{t+1} = bp_t + v_{t+1}$

$= -ba\, q_t + bu_t + v_{t+1}$

This is a recursive model in which p_t and q_t are determined in successive time periods by oscillations around stable long run functions.

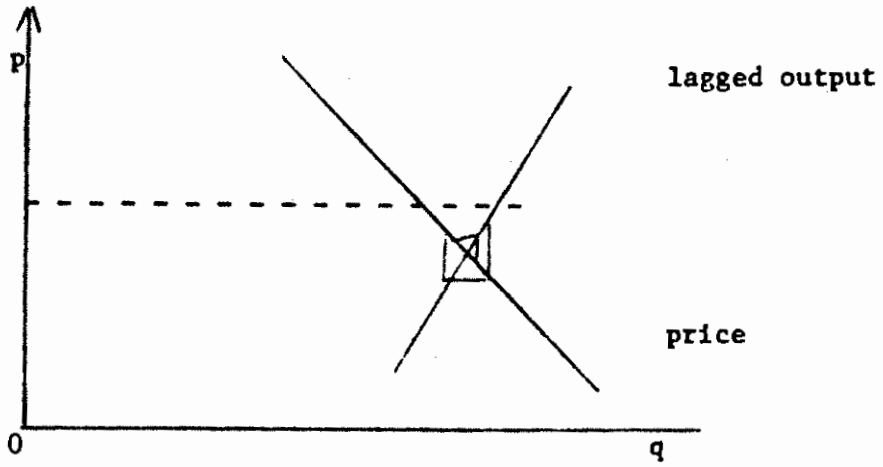


FIGURE 21.2

The product of parameters a and b determine whether the oscillations are convergent (ab < 1) or divergent (ab > 1).

Harlow[3/] used a cobweb model as a recursive system of equations for estimating quarterly hog supply and price functions. He noted that aggregation over time was an important consideration in the decision to use a recursive model. If a longer time unit such as one year were used it may permit simultaneous interaction of supply and demand. The resultant hypothesized simultaneity would result in biased estimates by OLS as the assumptions outlined earlier would be violated.

Harlow used OLS on quarterly data and found serial correlation is residuals. Non independence of residual terms resulted in loss of statistical efficiency and under-estimation of the variance. As a result t and F tests were no longer strictly valid. This serial correlation of disturbances resulted from exclusion of some variables, so Harlow included lagged quantity in the supply and demand functions.[4/] There was then no evidence of correlation among residuals of different equations and serial correlation was greatly reduced.

[3/] See Harlow, A. A., "Factors Affecting the Price and Supply of Hops", U. S. D. A. Technical Bulletin, 1274.

[4/] This was argued to be theoretically sound, and observationally verified by the tendency of economic variables to be serially correlated.

Harlow checked the condition of triangularity necessary for use of the recursive method. It is that the Jacobian of the transformation connecting the disturbances with the endogenous variables be triangular and equal to one.[5/] While not all of the off diagonal elements (covariances) were zero, this condition for use of OLS on a recursive system was approximately met.

The above examples suggest that recursive models are suitable for use in dynamic analysis situations, given that the econometric (derived) model can be correctly specified as recursive. Unfortunately the theoretical considerations discussed earlier are of little help in making the correct specification, the choice of recursive or simultaneous model depends on aggregation issues and the use of OLS versus simultaneous equation procedures.

Recursive and Simultaneous Systems Compared

As suggested earlier there are good reasons for assuming that the basic economic model is recursive. In the process of simplification and aggregation to develop an econometric model we may use either a simultaneous or a recursive approach. Which ever is decided on, the econometric model may be viewed as a set of hypotheses about a simplified representation of the real world. It is desirable to be able to specify the causality of the relations embedded in this set of hypotheses.

Recursive systems involve a direct specification of causality, as long as we move from equation to equation along the recursive chain suggested by Wold and others. In simultaneous systems, it is frequently true that no causal interpretation is intended in the structural model; the model merely hypothesizes that there are joint conditional probabilities of the endogenous and predetermined variables. This is a situation in which there may be

---

[5/] The Jacobian is the matrix of partial derivative of u_t with respect to Y_t.

causal circles, so that endogenous variables Y_1 and Y_2 are jointly determined. Such a situation makes sense only in an equilibrium system, which does not describe how equilibrium comes about, or the causal relations involved.

Strotz[6/] treats simultaneous models as either an approximation of a recursive model, or a description of its equilibrium state. In either case, he view simultaneity as a specification error due to the static nature of the model. This error may result from ignoring time lags and the dynamic adjustments in the model, or using "stock" rather than "flow" variables.

This view presented by Strotz is especially important in relation to aggregation over time in econometric models. In the particular case of Cobweb's models of hog supply and pricing, Harlow concluded that quarterly data would be recursive, but annual data would allow for simultaneity to occur. At the other extreme, when an individual consumer enters a store with a "conditioned" demand function, it becomes unconditioned as soon as the price is known; in this case recursivity could be hypothesized only for a flecting moment. Strotz generalizes the above examples by pointing out that lagged endogenous variables are considered current endogenous variables on the grounds that the lag is small relative to the observation period. This situation has some implications for model specification and for maximum likelihood properties of the estimated parameters.

The question of identification is of great importance in comparing recursive and simultaneous models. Recursive models have no identification problems, but in estimation of the parameters of a simultaneous structure, identification is an empirical problem of major importance. It has been argued that simultaneous models have estimation problems in anything but the over-identified case, because the model is a simplification of the real world

---

6/ For details, see Strotz R., "Interdependence as a Specification Error," Econometrica, Vol. 28, No. 2, April 1960.

and almost always omit some variables. This form of misspecification frequently results in underidentification. Fisher[7] advocates use of a recursive system as a way out of this problem. In a recursive set of equations misspecification in the form of omission of variables may cause problems in individual recursive equations, but the underidentifiability problem for the whole model, does not occur. The parameters of one equation in a recursive system may be biased by variable misspecification.[8] but other equations will not be affected. Further, Fisher quotes Wold's proximity theorems on the approximate fulfillment of OLS assumptions resulting in "nearly unbiased" coefficient estimates. This can be generalized to statements about the importance of misspecification of variables and its effect on simultaneous equation structural parameter estimates.

As a conclusion, Fisher seeks the middle ground of a "nearly" block recursive model. This seems to be explicit recognition of the possibility that choice between recursive and simultaneous models is a matter of judgment, and depends on the problem under consideration. As noted earlier, this choice will also depend on such factors as the degree of consumer, product, or time aggregation used.

In many respects simultaneous systems are more general in formal design than recursive systems. The latter have restrictive conditions on the matrix of endogenous variable coefficients, and on the matrix of residuals. On the other hand, recursive systems allow for a causal interpretation of behavior relations which can only be derived from the reduced form of a simultaneous

---

7/ Fisher, F. M., "On the Cost of Approximate Specification in Simultaneous Equation Estimation," Econometrica, Vol. 29, No. 2, April 1961.

8/ Variable mis-specification may take the form of omission of variables, incorrectly designating endogenous as exogenous variables, and vice versa, among others.

model. Wold[9/] attempts a generalization of recursive systems to combine the best features of both models into what he calls "conditional causal chain models."

Conditional causal chains are constructed on the basis of behavioral relations. Equilibrium conditions may then be imposed on the model to break the triangularity of the matrix of endogenous variable conefficients. The causal interpretation of the model is retained by using conditional expectations on residuals $E(u_t) = 0$. The objective of Wold's generalization is to distinguish between explicit (recursive) inference from the model and implicit (simultaneous) inference from the reduced form. Wold makes the following comparative comments.

1. Differences between the models are based on the mechanism hypothesized to generate the empirical data. It is therefore a theoretical rather than an empirical question. Elasticities can be interpreted directly from the recursive model but not from the structure of the simultaneous model. The notion of causal interprobability discussed earlier is important here.

2. In causal chains the disturbance is allowed for by predictive inference in the form of conditional expectations. In simultaneous models, you can specify relations as conditional expectations either in the reduced form or the structural form, but not in both forms concurrently. Thus causal chains have an advantage in simplicity of causal interpretation.

Recursive models are explicitly based on behavioral relations which Wold and others refer to as the "fundamental building material" of all econometric

---

9/ Op cit, Econometrica, Vol. 28, No. 2, April 1960.

models. The recursive model is explicitly dynamic in its approach to these relations, while simultaneous models impose static elements through the use of equilibrium conditions in the model.

Summary and Conclusion

In this the advantages of recursive and simultaneous models are briefly summarized and some of the main points of the controversy are emphasized.

1. Advantages of the Recursive Model.

   (a) Theoretical justification - using the rationale of Bentzel and Hansen leads one to a view of the real world at the micro level of behaving units, as being recursive. Whether or not this carries through to the econometric model as a simplified representation of the real world, is open to debate. In particular the aggregation and correspondence issues may lead to a simultaneous representation of a recursive world.

   (b) Ordinary least squares can be used for estimation. If the model is correctly specified, and the assumptions on disturbances hold, the desirable properties of BLUE estimators are present. In particular this results in robustness in testing the model. Attainment of such properties appears to be more of a problem in estimation techniques for simultaneous models.

(c) Direction of causality and behavioral interpretation are explicit products of recursive estimation. This is an advantage in forecasting and prediction. On the other hand, where no causal interpretation is desired, and the aim is to estimate structural parameters of a system, the simultaneous model may be equally applicable.

(d) Auto and serial correlation problems are not as great a problem in recursive models and there are no identification problems with recursive model estimation.

2. Advantage of Simultaneous Equation Models

(a) Given that the econometric model is a simplification of the real world with the associated aggregation and correspondence problems, simultaneity may be the most realistic approach to estimation.

(b) The simultaneous equation model is more general and flexible in that it is not restricted to any specific causal interpretation and emphasizes the joint determination of mutually interdependent variables. If the researcher is faced with questions of structure specifically, rather than prediction and forecasting, simultaneous models may be the most realistic approximation to the real world.

(c) Simultaneous systems involve less restrictive assumptions on the residuals in some respects. In particular the existence of correlation between the disturbances of different equations is permitted, and is in fact an essential feature of the simultaneity of the model.